GROUP TESTS OF INTELLIGENCE

GROUP TESTS OF INTELLIGENCE

BY
PHILIP BOSWOOD BALLARD
M.A., D.Lit.
AUTHOR OF "FUNDAMENTAL ARITHMETIC," "FUNDAMENTAL ENGLISH," ETC.

UNIVERSITY OF LONDON PRESS LTD.
WARWICK SQUARE, LONDON, E.C.4

First Edition 1922
Second Edition 1923
Sixteenth Impression 1957

Printed & Bound in England for the University of London Press Ltd.,
by Hazell Watson & Viney Ltd., Aylesbury and London

PREFACE

THERE is an old proverb which says that an ounce of mother wit is worth a pound of clergy, "wit" being the old English word for intelligence, and "clergy" the old English word for learning. And examiners have always been under the reproach of being unable to separate the wit from the clergy: they cannot tell us how much of a candidate's success at an examination is due to good brains and how much is due to good opportunities. It is the business of mental tests to remove this reproach. For it is the business of mental tests to extract the element of wit from the compound of wit and clergy. That they do their work imperfectly is readily admitted; but they are doing it better every day; and so important is the task that it is better to do it badly than not to do it at all.

There are people who think otherwise, people who believe that it is better not to do it at all. They say that mental testing is an attempt to bring exactitude into a realm which is essentially vague and indeterminate. They seek refuge in the plea that teaching is not a science, but an art; that the teacher is not a theorist dealing with abstractions, but a craftsman whose workshop is the classroom and whose material is the plastic mind of childhood.

But they find no real refuge there; for they are there exposed to the charge of refusing to learn about the material on which they exercise their craft. No other craftsman is so indifferent, or being indifferent would care to confess it. The weaver is curious about his staple and the potter about his clay; for the wares they turn out are in accord with the materials on which they work. And nobody would think of blaming the weaver who failed to make a silk purse out of a sow's ear, or the potter who failed to make a porcelain vase out of a lump of common clay. Indeed he would blame them for trying. And yet it is not so very long ago that the results of teaching little children were all measured by the same scale, and paid for at the same rate, in total disregard of those children's natural gifts. Silk purses and porcelain vases were expected everywhere. The parable of the talents was read on Sundays and forgotten on week-days. Applied to the classroom the parable means that the critic—even if that critic be the teacher himself or the pupil himself—cannot justly judge the fruit of the teacher's labours without taking into account the fertility of the field in which he works. And to discover the fertility of the field in which he works is the aim and purpose of mental tests.

In my book on Mental Tests I described certain methods of testing intelligence individually, and certain methods of testing attainments collectively. In this book I confine myself almost entirely to the testing of intelligence collectively—a mode of mental testing with which the teacher will feel quite at home; for it is akin to the ordinary written examination. Akin to it, but not identical with it.

The testing needs care, but does not need a special psychological training.

Certain readers of *Mental Tests* have expressed a wish that I should extend my chapter on statistical theory so as to include an account of correlation and probable error. That wish is met in this book.

My own contributions to the public stock of mental tests will be found in Chapters XII—XV. Teachers who wish to use these tests of mine are at perfect liberty to do so. Though it is possible for them, by a manifolding process, to prepare copies for themselves, they will find it more convenient, and scarcely more expensive, to purchase printed copies from the publishers of this book. The booklets differ from the American booklets in one important respect: they can be used over and over again.

The above remarks do not apply to the Northumberland Tests, which are the property of Messrs. George G. Harrap & Co., Ltd. From them may be purchased copies not only of these tests, but also of the following tests to which I have referred in this book: The Terman Group Test of Mental Ability, the Otis Group Intelligence Scale, the National Intelligence Tests, and the Simplex Group Intelligence Scale. I owe a debt of gratitude to these publishers for permission to quote freely from their tests.

I also acknowledge my indebtedness to Messrs. Sidgwick & Jackson of London, and to Messrs. Henry Holt & Co. of New York, for allowing me to cite examples from their excellent little book, *Mental Tests in the American Army*, by C. S. Yoakum and R. M. Yerkes.

My personal obligations make a formidable list. I offer my special thanks to Professor Godfrey H. Thomson for his kindness in allowing me to reproduce his Northumberland Tests; to Dr. C. S. Myers for permission to quote from two articles which appeared in the *British Journal of Psychology*, Professor Thomson's on the Northumberland Tests, and my own on the Limit of the Growth of Intelligence; to Dr. Myers again for kindly lending me a number of American tests in his possession; to Professor C. Spearman for a similar kindness; to Dr. William Garnett, his son Dr. J. C. Maxwell Garnett, and my colleague Captain J. Brown, for extremely helpful criticisms and suggestions on the chapters on Correlation and Probable Error; to Mr. E. A. Williamson for preparing some of the figures for the press, for improving on my rough drawings for the Picture Test and adding a few original drawings of his own; to my brother-in-law Mr. R. D. Bartlett for getting the diagrams ready for reproduction; to Dr. Leonard P. Ayres and Dr. W. Franklin Jones for permission to reproduce their lists of spellings; to the following teachers for help in standardising the tests: Messrs. J. G. Robson, A. E. Roberts, J. Scott, T. Lea, G. Jones, and W. A. Watson; and finally to my wife for correcting the proofs, except those of the two statistical chapters, where she begged to be excused on the plea that the subject-matter of those chapters could by no possibility engage the interest of a rational being.

P. B. BALLARD.

CHISWICK,
April, 1922.

CONTENTS

GROUP TESTS OF INTELLIGENCE

CHAPTER I

GENERAL CHARACTERISTICS

When the fate of a young child hangs upon the issue, or whenever there is need for punctilious care in making a diagnosis, the tests applied are always individual tests—individual and *viva voce*. They are always either the Binet tests, or a revision of the Binet tests; and the Binet scheme of testing is the old *viva voce* examination with a difference. The whirligig of time has here turned the most careless and casual form of examination into the most accurate and rigid. At the old-fashioned *viva voce* the examiner trusted to the inspiration of the moment. He did not prepare his questions, he did not standardise his material, he did not put the same tests to all the candidates. He had therefore no ground for the comparison of candidate with candidate; nor had he a scale of values by which he could collate the performances of the present with the performances of the past. Instead of taking a measure he formed an impression; and generally speaking his confidence in the impression was in inverse ratio to his skill as an examiner. But then it was never pretended that the *viva voce* was the real investigation: the real investigation

as the written examination, and to this the *viva oce*, when it was not a mere perfunctory piece of outine, was subsidiary and supplementary. In ne modern scheme of mental testing the two types ave changed places. The more searching and icisive inquiry is now oral and individual. But it as its drawbacks. It takes an unconscionable time, nd it demands on the part of the tester a knowledge oth of the technique of testing and of the principles nderlying the tests, such as is possessed by few. Jnder the individual system everybody cannot be ested, nor is it everybody who can test. Hence the eed for group tests.

Individual testing was born in France; group esting was born in America. And its mother was ecessity—the stern necessity of war. How the Americans tested their soldiers for the Great War s now an old story; but it is worth retelling. They hemselves, with pardonable pride, refer to it as 'the greatest piece of mental engineering the vorld has ever seen." They point to England's ailure to make the best use of her most valuable esource—the intellectual powers of her people. But then England had to act in a hurry, while America had ample time to think about things before she entered the war. When she did enter, one of her first steps was to use the psychology of ner psychologists. The claims made by the psychologists were by no means great. They said they could pick out the recruits who were not worth training —those who were so stupid that they could never make good soldiers. Even these modest claims, however, were viewed with scepticism by the General Staff; and the value of mental tests for

soldiers was put to a rigorous trial. A large number of men in the standing army were tested and classified by the psychologists, and this classification was compared with an independent classification made by officers who had known the men intimately for some time. The two classifications were compared; they tallied so well that the scepticism of the war authorities was broken down. It was frankly admitted that the psychologists could not only do what they predicted, but very much more. The upshot was that the scheme was enormously extended, both in scope and purpose. It was decided to apply it to the whole of the American Army with the exception of field and general officers; and it was officially proclaimed that the purpose of the psychological examinations was not only to eliminate the unfit, but also to assess the intelligence of every soldier in the army; to indicate the kind of training each recruit might most profitably undergo; to permit of the units being so organised that men of the same grade of inte ligence could serve together in the same regiment; and to pick out those men whose general ability marked them for promotion or whose special abilities fitted them for special kinds of military service. In other words, the scheme enabled the authorities to fire out the fools, to prevent the yoking together of ox and ass, and to put the right man in the right place. Altogether nearly one and three-quarter million men were examined, of whom 41,000 were officers.

The men were dealt with in large groups; sometimes as many as five hundred were tested at once. When, however, a man failed to score a reasonable

number of marks in the group test he was tested again by himself—tested on the Binet system. In fact he was not labelled mentally defective until it was proved beyond doubt that he deserved the label. Of the 83,000 men tested in this way, only one in ten proved to be so bad as to be utterly useless for any kind of military work.

The whole undertaking was a colossal business; and the official report which has recently been issued is correspondingly colossal. It weighs about four pounds, and is so congested with tables and diagrams and graphs, and regression equations and correlation coefficients, and correction formulæ and other fearsome things, that it is difficult to conceive anybody having the patience to read it through from cover to cover. To study it in full would "exhaust time and trench upon eternity." The important thing to know, however, is that the scheme succeeded—succeeded so well that the American nation is now so firmly convinced of the predictive value of group tests of intelligence that it is rapidly extending their use to nearly every department of life; especially to education. But let us return to the tests themselves.

In devising the tests two great difficulties had to be faced: how to make them "fool-proof," and how to apply them to illiterates. A fool-proof test is one that prevents the examiner from making a fool of himself. Each question is so constructed as to admit of one answer only—one clear, simple, unequivocal answer. To make quite sure, the examiner is told what that answer is. Nothing being left to his individual judgment, he cannot possibly go astray. Such a test is fool-proof. Let

me illustrate. Binet has placed in the tenth year of his scale the following test :

"If you were asked what you thought of someone whom you did not know very well, what should you say ?"

Binet's instructions are that any answer is to be regarded as satisfactory if it suggests the need either of making an inquiry or of withholding an opinion.

The same question, with the words "very well" left out, appears in the American Army Tests ; but it is followed, as below, by three suggested answers, and the examinee is asked to make a cross in the square that stands before the best answer.

☐ I will go and get acquainted.
☐ I think he is all right.
☐ I don't know him and can't say.

Since the key indicates the third as the correct answer, the examiner is not free to consider either of the others. If, however, he were acting on Binet's instructions he would hesitate to reject the first answer ; he might accept it, or he might not. In fact Binet's question is not fool-proof : the army question is. All the army questions were of this type ; each demanded as definite and specific a solution as the answer to the question : How many twos are contained in ten ?

The device of alternative answers, though first extensively used during the war, was previously employed by Mr. W. H. Winch, as is shown by this typical example of his reasoning tests of 1912 [1] :

"Tom had twice as many mistakes in dictation as Harry, and Jim had twice as many as Harry. Did Tom have more mistakes than Jim, or less than Jim, or the same number, or can't you tell ? You must say why."

[1] *British Journal of Psychology*, vii, 190–226.

The army group tests to which I have so far referred are those which were applied to the bulk of the American Army and were known as the Alpha tests. They presupposed the ability to read and to understand the English language. But nearly 30 per cent. of the soldiers could not do both of these things, either because they were completely illiterate, or because they were foreigners who had not yet mastered the language of their adopted country. To these the Beta tests were given, a distinct set of tests consisting of figures, diagrams, pictures, and puzzles—tests couched in the universal language of line and form, and of so obvious a nature that it was possible, with the help of a blackboard and a few gestures, to explain clearly what was required to be done.

I have dealt at some length with the American Army Tests because they are not only important in themselves, but important as the parents of a large progeny of group tests which are still multiplying. It is true that there were group tests before the days of Yerkes and his committee, just as it is true that there were individual tests before the days of Alfred Binet; but as Binet put us on the right track for individual testing, so has the American committee put us on the right track for group testing. At any rate it is the track that has been followed ever since. It is the track that has been followed by the two or three dozen standardised group tests that have already been printed and published, and circulated in the American schools. The more popular of them have found their way to England and are issued in an Anglicised form by Harrap & Co. If there is wisdom as well as

safety in a multitude of counsellors, the highest value must be attached to the series prepared under the auspices of the National Research Council, and known as the National Intelligence Tests. The committee of psychologists responsible for them is virtually the same committee as were responsible for the war tests. And the tests cost 25,000 dollars to prepare. But in spite of the name (and the cost) they do not seem to be better, or more extensively used, than others of more private origin, such as the Terman Group Tests, the Haggerty Group Tests, and the Otis Group Tests. The Otis Group Tests are of special interest, as some of them were in manuscript before the war, and seem to have formed the original nucleus of the Army Tests.

One marked characteristic of all these group tests is that they have to be worked at break-neck speed, as will be seen from an inspection of Tables I and II. At the Alpha examination 212 questions had to be answered in 23 minutes 15 seconds, which makes an average of about 6½ seconds per question.

TABLE I

ARMY GROUP EXAMINATION ALPHA

No. of Test.	Name of Test.	No. of Questions.	Time Allowed.	
			min.	sec.
1	Instructions . .	12	2	15
2	Arithmetical Problems	20	5	0
3	Best Answer . .	16	1	30
4	Word Meaning . .	40	1	30
5	Mixed Sentences .	24	2	0
6	Number Series . .	20	3	0
7	Analogies . . .	40	4	0
8	Information . .	40	4	0
	Total . .	212	23	15

The whole examination of course takes much more time than is indicated in the table; in fact it takes about 50 minutes, because there are pauses between the tests, and explanations of the test that is to follow.

In the tests intended for schools the time allowed is a little more liberal. Terman, for instance, as will be seen from the analysis in Table II, allows as an average about $8\frac{3}{4}$ seconds for each question.

TABLE II

THE TERMAN GROUP TESTS

No. of Test.	Name of Test.	No. of Questions.	Time Allowed.
			min.
1	Information	20	2
2	Best Answer	11	2
3	Word Meaning	30	2
4	Logical Selection	20	3
5	Arithmetic	12	4
6	Sentence Meaning	24	2
7	Analogies	20	2
8	Mixed Sentences	18	3
9	Classification	18	3
10	Number Series	12	4
		185	27

But even this means working at the rate of seven questions per minute—a pace which, even though most of the questions are easy and no time is spent in writing, seems to the plain man to be a wanton incitement to guessing and a wicked discouragement of thinking. Indeed the most serious criticism that has been made against the American group tests is

that they put a premium on smartness—that they pick out the rapid thinkers and leave behind the profound thinkers. Those who devised the tests look upon brain-power just as engineers look upon horse-power: they regard it as a thing to be measured by the amount of work it can do in a given time. And this indeed is inevitable if we consider intelligence as including the ability to deal expeditiously with certain common tasks. Even Binet sets time limits to some of his tests. For instance, in his counting test for eight-year-old's (" count backwards from 20 to 1 ") he allows only 20 seconds. He gives a child of twelve only one minute to rearrange the mixed sentence, " a defends master dog good bravely his." It is clear that if unlimited time were allowed, such questions would lose in distributive and diagnostic value. The valid objection is not that some of the army tests have time limits, but that all of them have time limits —that they contain no tests at all which give an equitable chance to the slow, cautious, and solid thinker. It is to meet this objection that in my own group tests (see Chapters XII—XV) some, if not all, of the questions are to be worked at the candidate's own pace.

Another characteristic of the American group tests is that the answers take the form of marks placed on the question paper itself. A word has to be underlined or crossed out, a number has to be inserted, or a sentence has to be indicated by a cross. So numerous are the questions that each examinee has to be provided with a booklet, each page of which presents one type of test. Here we find a striking contrast with the ordinary school

examination. In an ordinary school examination the questions are, comparatively speaking, short and the answers long; in these tests the questions are long and the answers short—so short and unequivocal that they can be marked by stencil. The stencil consists of a sheet of transparent celluloid, or of tough tracing paper, which exactly fits the page of the booklet and has printed on it in appropriate places the correct answers. By superimposing the stencil on the page all errors can be immediately detected. When, however, as often happens, the answers range in a vertical column on the right side of the sheet, it is just as convenient to use a piece of cardboard which has the correct answers printed at the corresponding intervals along the edge.

Yet another characteristic of the tests is the prefatory sample. The eighth test, for instance, of the Alpha examination begins thus:

> Notice the sample sentence:
>
> People **hear** with the **eyes ears nose mouth.**
>
> The correct word is ears, because it makes the truest sentence.
>
> In each of the sentences below you have four choices for the last word. Only one of them is correct. In each sentence, draw a line under the one of these four words which makes the truest sentence. If you cannot be sure, guess. The two samples are already marked as they should be.
>
> SAMPLES:
>
> People **hear** with the **eyes ears nose mouth.**
> **France** is in **Europe Asia Africa Australia.**

Then follow forty similar sentences, for dealing with which four minutes are allowed. Every other test, not only in the army series but also in school series, is introduced in the same way. Some indeed go

still further and provide preliminary exercises as well as preliminary samples: they give the examinee sufficient practice in the type to ensure that in the actual test he is not impeded by doubt as to what is required of him. Such practice series appear in the National Intelligence Tests and in the Haggerty Tests—practice series which are about a third as long as the test series.

Nearly all group tests are published in at least two sets, which are supposed to be of equal difficulty, which carry precisely the same marks, and which are paradoxically called forms—paradoxically because the form of the tests is the same: it is only the content that is different. The first question, for instance, in Form 5, test 2, of the Alpha examination runs thus:

How many are 30 men and 7 men?

The first in Form 6, test 2, is:

How many are 40 guns and 6 guns?

These two questions do not only appear on *a priori* grounds to be of equal difficulty: they also proved in actual practice to be of equal difficulty. Thus every item in one form is matched by a corresponding item on the alternative form. In the American Army Tests there were five such alternative forms, numbered 5 to 9. The National Research Council have already issued two alternative forms of the National Intelligence tests and have promised to issue others from time to time. The object of issuing alternatives is to prevent coaching, and to provide tests for candidates already familiar with the original set of questions.

It will be seen that in these intelligence examinations tradition is persistently set at naught. The "ten minutes' puzzles," as De Morgan used to call the examination questions in mathematics, are replaced by the briefest of problems in mental arithmetic. The essay, on which the examiner has always set so great a store, has passed away and left no trace behind. And by the essay I do not mean merely the theme, the thesis, or the set exercise in composition, but the answers required to the questions in history, geography, science, and other branches of learning. For each of these answers is a little essay in itself, with all those complexities of the essay, all those dubious points of merit and of demerit, which render it utterly unsuitable as a means of exact measurement. For an essay can never be made fool-proof. Even the Americans, with all their genius for standardising, have failed here. None of their composition scales are genuine scales; nor do their valuations in any way convince us. Indeed their notion of what constitutes good English often seems to us to be strange and perverse. One measuring scale for English composition has, for instance, only five grades of merit. The sample of the highest grade—a sample labelled "Superior Quality"—contains the following sentences: "Bobby said they came up close so we could hear if anything got them." "I am rather doubtful as to whether I will be granted a vacation this year." On the whole we think the Americans have acted wisely in keeping composition out of their intelligence examinations.

The long answer does not suit the group tester, for he always aims at fine calibration. Each group

scale is an aggregate of little things; each test contains a number of little questions, all of the same type, questions which sometimes differ from one another in difficulty almost as little as one millimetre differs from another millimetre in magnitude.

In an ordinary examination it is the answers that absorb the time and the energy of the examiner. He often throws off the questions with a fine carelessness, and pays for it dearly in blue pencil and midnight oil. In group testing, on the other hand, the whole mind and strength of the examiner is put forth in devising the questions. The answers trouble him not. There is no need; for they may be marked by a clerk, or a schoolboy, or possibly, in the near future, by a machine.

CHAPTER II

INSTRUCTIONS TESTS

THE first test in the Binet Scale is an instructions test, and the first test in the Alpha scale is an instructions test. The examinee is asked to do something quite simple—to point to his eyes, nose, and mouth, in the Binet examination, and in the army examination to make marks on a sheet of figures, diagrams, and words. On the first line of the sheet there are five circles in a row, and the first instruction is given orally thus :

"Attention! 'Attention' always means 'Pencils up.' Look at the circles at 1. When I say 'go,' but not before, make a cross in the first circle and also a figure 1 in the third circle.—Go!" (Allow not over 5 seconds.)

On the twelfth and last line of the sheet are printed the natural numbers from 1 to 9, and the instructions are as follows :

"Attention! Look at 12. If 7 is more than 5, then (when I say 'go') cross out the number 6 unless 6 is more than 8, in which case draw a line *under* the number 7.—Go!" (Allow not over 10 seconds.)

All these instructions are given by word of mouth and are uttered once only; so that if the candidate does not catch the meaning at once he is lost. To be quick in the uptake, to be able to obey orders readily,

is obviously a valuable gift, not only in soldiering, but in nearly every circumstance of life. At any rate one feels it to be a valuable gift for others to possess. The mother makes heavy demands upon it in the home and the teachers in the school. Mr. Wackford Squeers was not unaware of its advantages. But in these cases it is exercised rather than tested: the ability is supposed to be there in unlimited measure. One feels it approaching the nature of a test when one asks a policeman in the street the way to such and such a place, and gets as a reply: "Take the first turning to the left, then the second to the right, then the first to the right again; keep straight on till you get to the fire-station and turn down the street just beyond the fire-station which has a public house at the corner, and so on, and so on." Personally I always fail in this test.

The instructions test is obviously a test of listening, of understanding, and of remembering. The testee must listen carefully, he must understand clearly, and he must bear in mind what he is told until the order is completely carried out. Sometimes, the instructions being presented in writing or in print, the listening is eliminated, and the test is rendered easier by the fact that the testee may reread the directions as often as he likes—within the time limit. But it is quite possible to make up for this by complicating the instructions a little more. The usual way is to impose certain conditions and reservations—to add a judicious sprinkling of "if" clauses and "unless" clauses.

In adult life we meet the printed instructions test—or rather it meets us: we do not seek it—in the guise of official forms and *questionnaires*. And

we acquit ourselves badly. Administrative officers assure us that about 70 per cent. of the returns sent in by educated citizens are wrongly filled up. And they say unkind things about our intelligence. But it is our patience that is at fault, not our intelligence. For these forms are far duller than the American group forms, and their impertinence and inquisitiveness (I have specially in mind the forms that are headed "Income-tax") compare badly with the fine impersonality of an examination paper.

The Otis advanced examination leads off with a good instructions test. At the top of the paper the Alphabet is printed in capitals, and each of the twenty questions on the page bears upon this alphabet. I will quote three:

5. Write the letter O.

14. Write the letter which is the third letter to the right of the letter which is midway between K and O.

20. Find the two letters in the word AFTER which have just as many letters between them in the alphabet as in the word. Write the one of these letters that comes first in the alphabet.

In reading these it will be seen how easy it is for the instructions test to pass into a sort of puzzle, where the difficulty is not so much in remembering what one is told to do and then doing it, as in finding out by an elaborate mental analysis what it is that one is told.

The instructions type of test is frequently used nowadays in our infant schools as a reading test—as a test of the child's capacity to absorb the meaning of what he reads. A teacher, for instance, will give a child a card on which is printed: "When you have read this, turn the card upside down on the desk, go to the blackboard and write the word 'Yes';

then come to me and say 'rabbits' "—or some such rigmarole. Between it and the American instructions test there is no difference except in purpose. Indeed the questions in one of the most popular reading tests in America (the Kansas Silent Reading Tests) are often indistinguishable from instructions. Test I, No. 5, for example, runs thus:

If you would rather have a dollar than a little stone, do not put a line under dollar; but if you would rather have five dollars than a pencil, put a line under stone.

dollar stone

Among the "Do as it says" tests given by Trabue in *Measure your Mind* the following may be quoted as a curiosity:

If ontogeny invariably ingeminates phylogeny, circumscribe the word giving the location of the OURCQ; if not, underscore the word that locates the MANDIBLE.

England Foot Utah Face Peru France Arm India

This delightful piece of nonsense would have warmed the heart of Lewis Carroll. To ingeminate phylogeny sounds like some unholy deed to be done at dead of night with masks and dark lanterns; and as for the word that locates the mandible, it surely is not to be found in the vulgar list that follows; it belongs to the same realm of mystery and romance as the name of the youth who slew the jabberwock.

CHAPTER III

ANALOGIES

IT is recorded by Boswell that a certain Dr. Adams called one day upon Dr. Johnson and found him busy at his dictionary. The following dialogue ensued:

ADAMS: But, Sir, how can you do this in three years?

JOHNSON: Sir, I have no doubt I can do it in three years.

ADAMS: But the French Academy, which consists of forty members, took forty years to compile their Dictionary.

JOHNSON: Sir, thus it is. This is the proportion. Let me see; forty times forty is sixteen hundred. As three is to sixteen hundred, so is the proportion of an Englishman to a Frenchman.

If Johnson had been a mathematician instead of a man of letters he would, to express himself more precisely, have inverted the ratio. But his meaning is clear. This attempt to humanise mathematics and to bring the method of proportion into the ordinary affairs of life is an anticipation of the modern analogies test, which is nothing but our old friend the "rule of three" with words taking the place of numbers. Three terms are given, and a fourth has to be found bearing the same relation to the third as the second bears to the first.

Analogies were first used as intelligence tests by Mr. Cyril Burt in his investigations at Liverpool in 1910. They have since become very popular. They have been inserted in some of the revisions of the

Binet Scale, and it is rare to find a group scale which does not include analogies as one of its tests. Of the hundred analogies which Mr. Burt gives in *Mental and Scholastic Tests* the first five are as follows:

1. Prince *is to* Princess *as* King *is to* —— ?
2. Pencil *is to* Drawing *as* Brush *is to* —— ?
3. January *is to* February *as* First *is to* —— ?
4. Sailor *is to* Soldier *as* Navy *is to* —— ?
5. Moon *is to* Earth *as* Earth *is to* —— ?

Here there is no ambiguity about the question : no uncertainty about the answer. Each problem has but one possible solution : each is virtually fool-proof. But all analogies are not so. This one, taken from another source, admits of several answers :

Bantam *is to* Fowl *as* Merino *is to* —— ?

Any of these words will fit—beast, animal, sheep, wool, fabric, cloth—and possibly a few more.

The Alpha Tests remove the possibility of variant answers by appearing in the following form (I give the first two and the last two of a series) :

1 gun—shoots : : knife—**run cuts hat bird**
2 ear—hear : : eye—**table hand see play**
39 tears—sorrow : : laughter—**joy smile girls grin**
40 cold—ice : : heat—**lightning warm steam coat**

The examinee is asked to choose the appropriate word from among the **four in heavy type,** and to underline it. But the test now seems to be made fool-proof at the expense of robbing it of the greater part of its value as an index of thought. For it is much easier to select one term out of four possible terms than to think out the relationship without any hint or help of any kind. Besides there is no clear proof that the testee has really

grasped the relationship; for if he selected one of the words at random he would be right in about 25 per cent. of the cases. The reply made to this criticism is that among the words presented there is always at least one that tempts the testee to go astray—one that is more usually associated with the third term than is the correct answer, and therefore, to the unthinking, seems more plausible. That this plea is fully justified in the army examples I gravely doubt. In the first two examples cited above, where is the association link that is stronger than the one that establishes the logical relationship? The mere impetus of habit carries one here along the right path.

But there is another, and a more conclusive, way of comparing the two methods—the way of experiment. Children have been tested with analogies where the choice of the fourth term is left free and unfettered; and again with analogies where the possibilities are narrowed down to four. And the difference in the two classes of results is regarded, by the Americans at least, as insufficient to outweigh the advantage of the fool-proof key.

It sometimes happens that when the analogy is controlled one is forced to give a different solution from what one would give if the analogy were free. Terman, for instance, includes the following among his group tests:

4 is to 16 as 5 is to —— ?

The reader with a mathematical training would at once say 20, were he permitted to do so. But he is not permitted, for he has to choose his answer from among these four numbers: 7, 45, 35, 25.

The missing term is almost invariably the last term. In the Miller Mental Ability Test, however, the position of the missing term is constantly changed. For example :

FISH : SWIM : : (man, boat, flower, tree) : WALK.
(hour, week, long, day) : NIGHT : : WHITE : BLACK.
TOE : (knee, foot, arm, nail) : : FINGER : HAND.

The most difficult analogies I have come across appear in Roback's Mentality Tests for Superior Adults. He puts his sample in this form :

Sun—heat ; flower—odour.

and leaves the subject free to choose his own fourth term, as in the following :

At—that ; is—
/—\ ; ×—
Square—triangle ; circle—
Mate—fate ; book—
Except—accept ; extend—
8—o ; ×—
9—6 ; p—

Analogies appear in a new guise in Thurstone's Psychological Examination for College Freshmen and High School Seniors, as will be seen from the following specimens :

Underline two words with the same relation as egg and bird.

Crack seed plant grow nest.

Underline two words with the same relation as boy and man.

Wool lamb dog sheep shepherd.

There can be no doubt as to the importance of the mental factor the analogies test is designed to measure. William James in a well-known passage [1] says :

" Geniuses are, by common consent, considered

[1] *Principles of Psychology*, vol. ii, p. 348.

to differ from ordinary minds by an unusual development of association by similarity. One of Professor Bain's best strokes of work is the exhibition of this truth. It applies to geniuses in the line of reasoning as well as in other lines. And as the genius is to the vulgarian, so the vulgar human mind is to the intelligence of a brute. . . . The most elementary single difference between the human mind and that of brutes lies in this deficiency on the brute's part to associate ideas by similarity."

To associate ideas by similarity may therefore be regarded as one of the cardinal characteristics of the mind of man, and of all the ideas we associate by similarity ideas of relationship are probably the most important. For ideas of relationship underlie all our reasoning processes, they form the main staple of mathematical thought, and they find abundant expression in the literatures of all nations. Curiously enough William James, though he probably never heard of the analogies test, uses in the very passage quoted above as complete an analogy as any to be found in the American Army examination. Overtly or covertly it appears everywhere; in the most frivolous of jokes no less than in the sublimest of poetry. Indeed the fashionable joke of the hour hangs upon a specious analogy. A graceless youth (or maiden) springs upon you the question: "Have you seen Lydia?" And if you are unwary enough to ask "Lydia what?" you get as a reply, "Lid o' your dustbin," or some such piece of nonsense. Your mind following one line of relationship is suddenly jerked on to another. And the jolt amuses you—if you are in the mood for it.

CHAPTER IV

THE MEANING OF WORDS

ONE of the most persistent charges brought against the Binet Scale, and indeed against every other general scheme of testing intelligence, is that it attaches too much importance to words and too little to things. It is biased towards language. It does not give a fair chance to the child with a strong practical bent—the child who is inarticulate in speech but has a peculiar gift of thinking, as it were, through his fingers. This criticism is doubtless well grounded : it points to a real defect in the mental tester's equipment. But when all this is admitted the fact remains that if there is one universal type of mental ability it is the linguistic type. Even the practical man, practical in the narrow sense, is not so inarticulate as he seems : he is articulate enough in the matters that interest him. And every branch of science, however practical its bearings, is, as has well been said, nothing but a well-constructed language. Language is so pre-eminently the instrument of thought that the active mind cannot help using it. An intelligent child of two cannot help knowing the meaning of the word " bread " ; nor can the intelligent child of seven help knowing the meaning of the word " book."

It is therefore contended by some that if we measure a child's vocabulary we go a long way towards measuring his mind. Terman holds this view, and includes in his Stanford Revision of the Binet-Simon Scale a vocabulary test of his own, of which he makes wide use. The child is asked to define as many words as possible from a long list. Binet himself had a definitions test, but he limited his list to five simple words (fork, table, chair, horse, mother) and based his diagnosis mainly on the type of definition used. Terman, on the other hand, has a list of a hundred words of increasing difficulty, and he is not particular about the precise mode of definition so long as the child gives clear evidence that he knows what the word means. And he has a theory that we can roughly gauge the full range of a child's vocabulary by counting the number of test words with which he is familiar and multiplying it by a given constant. But this test is oral and individual.

For group testing, Mr. Burt's definitions test[1] is probably the best. It is so constructed as to afford a clue not only to the general ability of the person tested, but also to his particular bent of mind. The test, however, is not easy to mark. Indeed no fool-proof definitions test seems to have appeared yet, though such a test would not be difficult to devise.

Considerably the most frequent form of word-meaning test is the opposites test. A list of words is presented to the child and he is asked to write down against each of them the word that means

[1] *Mental and Scholastic Tests*, pp. 229–30.

the opposite. Here we have a valuable test of vocabulary; for it secures evidence of knowledge of two terms at once, each of which illumines and defines the meaning of the other. When I hear a speaker or writer use a term of vague signification I often find it useful to discover what he has in the background of his mind as the opposite term. The word "imagination," for instance, as used by the man of letters, generally means mental activity not tied down to the things of sense—ideation, in fact. Its opposite is sense-perception. When an artist exhorts a student to cultivate his imagination, he means thought embodied in concrete mental imagery as distinct from thought accompanied by symbolic imagery or by no imagery at all. The opposite here is conception. Professor Stout again brings imagination and belief into sharp opposition. Whatever value these may be in this habit of thought, there is a fair consensus of opinion that opposites afford valuable material for testing general ability.

Otis in his group tests presents opposites in foolproof form. I here cite his prefatory samples:

up . . . (short, down, small, low, young)
hot . . . (warm, ice, dark, cold, fire)

The complementary test—the synonyms test—is less frequently used. Burt, however, has employed it and found value in it, though not perhaps as much value as in his opposites test.[1]

The usual course adopted in modern group tests is to combine these two types in a synonym-antonym

[1] Opus cit. 225–6, 228–9.

test, as was done in Test 4 of the Alpha Examination. The test was introduced thus :

If two words of a pair mean the same or nearly the same, draw a line under *same*. If they mean the opposite or nearly the opposite, draw a line under *opposite*. If you cannot be sure, guess. The two samples are already marked as they should be.

SAMPLES :

good—bad.....same—opposite
little—small.....same—opposite

Then followed forty pairs of words, each aligned with the pair, same—opposite. It is clear that this makes the test fool-proof, but it is by no means clear that the statistical results are sound. There are only two alternatives to choose from, and if the examinee makes his choice by tossing up a coin he will as a rule get half his answers right. It is claimed that this difficulty has been met by counting as the real score not the number right, but the number right minus the number wrong. If, therefore, the candidate attempts the whole forty and gets twenty right his score will be zero. This device of marking seems to meet the difficulty, it is true ; but many cannot get rid of an uncomfortable feeling that there is still too much scope for guesswork, especially as the examinee is told to guess if he does not know.

It is a good feature of the test that the words get progressively harder as they get lower down the series. The last five pairs in form 6 are these :

36. suavity—asperity
37. perfunctory—meticulous
38. lugubrious—maudlin
39. desuetude—disuse
40. adventitious—accidental.

Many other group tests have followed the army tests in this matter, and a synonym-antonym test, similarly presented, appears in the National Intelligence Tests, The Otis Primary Examinations, The Terman Group Tests, and the Haggerty Intelligence Examination Delta 1. It is worthy of note that the two group tests that have special examinations for juniors (the Otis and the Haggerty) put the test in the junior paper and omit it from the senior.

There is one form of word-meaning test which merges into the sentence-meaning test. It is what is known to the learned as the Ebbinghaus Completion Test, and to the public as the missing-word competition. By including it in the National Intelligence Tests the authors renounce the advantage of *le mot juste*—the one inevitable word—and make it very troublesome to mark. One of the sentences runs thus :

> Power is....from steam, petrol, ... and other things.

According to the key 14 different words are permissible for the first blank and 12 for the second. In the 20 sentences given there are 30 blank spaces, and only 3 of these can be fitted with one word alone.

Dr. Otis succeeds better. This is how he manages it :

NARRATIVE COMPLETION

DIRECTIONS.—For each numbered blank in the story, choose the best word of the three in the list having the same number as the blank. Underline the word you choose. You may write these

words in the blank spaces if you wish, but only the underlining counts. Do nothing about the blanks that are not numbered.

The Reward of Kindness	*Underline Words here*
Once upon a (1) there was a (2)	1 time place man 1
that lived in a (3) One as he was	2 man lion dog 2
roaming about, he stepped on a and	3 street garden forest 3
it stuck in his (5) (4)	4 tack thorn rock 4

The story continues thus for 32 lines.

CHAPTER V

THE MEANING OF SENTENCES

THERE is one point on which books on school method are unanimous and emphatic: they all warn the teacher against putting questions which can be answered by Yes or No. The implication is that when he gets his answer he can only be half sure that it testifies to genuine knowledge. This piece of advice, whether it be good or bad, is flagrantly violated in the American group tests. Two out of the eight Alpha Tests, embracing altogether 64 questions, are cast in this mould; and the proportion is not very different in the other published scales. The questions most liable to appear in the Yes or No form are those which are designed to test a knowledge of meanings. One of the National Intelligence Tests contains 40 questions, of which I will quote five:

1.	Do children play?	Yes	No
16.	Are ducks covered with fur?	Yes	No
24.	Do convicts usually preach sermons? . . .	Yes	No
35.	Do transparent goggles transmit light? . .	Yes	No
36.	May a titanic bulk be called infinitesimal? . .	Yes	No

The examinee is required to underline the right answer to each question. It will be seen that the questions are of increasing difficulty. Terman's

questions are fewer, more difficult, and more uniform. He begins with ·

1. Does a conscientious person ever make mistakes ? . Yes No

He ends with :

24. Is rancour usually characterised by persistence ? . Yes No

A more usual way of treating sentence-meanings is to use an adaptation of Binet's mixed sentences test. The examinee is presented with a jumble of words which he is asked to straighten out into a sensible sentence and to indicate whether what it says is true or false by drawing a line under one of the two words at the end of the line. The following are extracted from the Alpha Tests :

1. lions strong are . . . . . true false
6. gotten sea-water sugar is from . . . true false
15. not eat gunpowder to good is . . . . true false
24. repeated call human for courtesies associations . true false

In all these "hit or miss" cases the score is counted by subtracting the number wrong from the number right. But as I pointed out in the last chapter even this stringent system of marking does not fully satisfy those who think there is something essentially unsound in narrowing down the answer to two alternatives. It must be remembered, however, that the results of both the army tests and the National Intelligence Tests have been submitted to the most rigid statistical analysis, and it is difficult to believe that a fundamental defect in the mode of presenting the tests would have escaped detection. A very good defence of this type of question is made out by Mr. Arthur I. Gates.[1] As a means of estimating progress

[1] *Journal of Educational Psychology*, xii, pp. 276–87.

in college studies he regards the true-false test as far superior to the ordinary examination—the "essay examination" as he calls it. I quote some of his questions:

(*a*) The auditory area is in the temporal lobe of the cortex.
(*b*) According to Terman's findings, the average intelligence of boys is markedly superior to the intelligence of girls.
(*c*) The curve of forgetting has been more thoroughly investigated than the curve of learning.
(*d*) If we find a large percentage of mentally inferior children in the slums, it proves that slum environment is one cause of mental deficiency.

All the student has to do is to place opposite the question the word "true" or the word "false." We have no difficulty therefore in believing Mr. Gates when he assures us that the system is more popular than the essay examination both among the students and among the examiners. But he goes further, and gives statistical evidence to show that it is also more highly correlated with intelligence, and has greater predictive value.

Another method of dealing with the disarranged sentence is to be found in a very ingenious series of tests known as the Pressey Cross-out Tests.[1] There are only four tests, and the first of them is introduced by these three examples:

(*a*) see a I man on.
(*b*) knife chair the sharp is.
(*c*) John broken window trees has the.

Each example can be made into a sensible sentence provided one word is crossed out. There are 25 such sentences altogether, and the examinee is asked to cross out in each of them the superfluous word.

[1] *Journal of Educational Psychology*, xi, pp. 91–100.

There are two other ways of dealing with the mixed sentence: one is to ask the examinee to mark the first word of the reconstructed sentence, and the other is to ask him to mark the last word. I regard the latter as the better method, for it alone gives clear evidence that the words of the sentence have been set in order from beginning to end.

Otis includes in his advanced examination an interesting Proverbs test. A list of 12 proverbs is given, followed by a list of 12 statements, each of which explains one of the proverbs in the first list. The two lists follow a different order. The problem is to place in the brackets that stand before each proverb the number of the statement that explains it. In other words, the testee has to fit the right explanation to the right proverb. His first two proverbs are:

( ) Make hay while the sun shines.
( ) A drowning man will grasp at straws.

The explanations that fit are respectively:

(3) Make the best of your opportunities.
(6) Desperate people cling to absurd hopes.

In Whipple's Group Tests for Grammar Grades the subject is asked to match English Proverbs with African Proverbs by placing in the brackets the appropriate number. I quote these examples:

English Proverbs.	*African Proverbs*
( ) Married in haste we repent at leisure.	4. Distant firewood is good firewood.
( ) Distance lends enchantment to the view.	5. Ashes fly in the face of him who throws them.
( ) Curses come home to roost.	11. Quick loving a woman means quick not loving a woman.

Thurstone's Psychological Examination for College Freshmen and High School Seniors contains proverbs tests of the following sort :

Don't put all your eggs in one basket.

Check two of the following statements with nearly the same meaning as the above proverb :

....The mouse that has but one hole is soon caught
....Catch the bear before you sell his skin.
....The proof of the pudding is in the eating.
....Put not all your crocks on one shelf.

CHAPTER VI

ARITHMETIC AND LOGIC

THOSE who think there is a clear line of cleavage between tests of intelligence and tests of school attainments will be surprised to find that there is no set of group tests of any repute which does not contain at least one arithmetic paper. The first written test in the Alpha examination is an arithmetic test. The first of the National Intelligence Tests is an arithmetic test; and as this is wholly concerned with problems, another test is included which consists of abstract arithmetical examples of a purely mechanical type. A third test is devoted mainly to the checking of duplicate numbers. These arithmetic tests need not be described in detail, for they do not differ in essentials from the ordinary arithmetic examination. The examples are, however, simpler; they can nearly always be worked in one's head; and they take up much less time. If the candidate finds it necessary to make written calculations, he has to make them on the exiguous margin of his question paper.

The arithmetic paper is, with hardly an exception, a test in arithmetical reasoning rather than a test in computation—which is what one would naturally expect in an examination of intelligence. What first strikes the man who looks with fresh eyes at modern group tests is the rareness of any direct

and obvious appeal to logic. Most of the response called forth seems to be of an ordinary routine kind, involving, it is true, the exercise of common sense and of most uncommon alertness, but making no demands upon that ingenuity which solves in a third line a two-lined puzzle beginning: "Not all things that have handles are brooms." The syllogism is palpably absent. Barbara, Celarent, and the rest of them are entirely ignored. Even inductive reasoning does not figure conspicuously. And yet it is not only a popular belief, but a scientific hypothesis which has received no small amount of experimental support, that the nearer we approach reason the nearer do we get to the very soul and secret of intelligence.

A closer inspection of the tests, however, reveals the fact that the reasoning is there, though it is not manifest on the surface. All the elements of logic as they operate in the work-a-day world, if not as they appear in the pages of Aristotle, are called into play by these simple-looking tests. Let us consider, for instance, the following examples from test 6 of the Alpha examination:

Directions: Look at each row of numbers below, and on the two dotted lines write the two numbers that should come next.

2	3	4	5	6	7	....	....
9	1	7	1	5	1	....	....
15	16	14	17	13	18	....	....
3	4	6	9	13	18	....	....

In the first of these examples it seems as though the mind's own momentum must inevitably carry it forward the next few steps; yet this momentum itself is due to a previous process of learning, and perhaps of analysis and synthesis. As the series get

more difficult, the analysis and synthesis, the investigating and the inferring, more obviously take place in the present; and in such examples as the last two above, one with its double pattern and the other with its single pattern, all the processes of inductive logic—inspection, hypothesis, and verification—have wittingly or unwittingly to be gone through before the law of the series is confidently apprehended.

This number-series test is a popular one, and appears in other forms. Trabue leaves the blanks in the middle of the series instead of the end, thus:

Series 3	.	.	12	11	10	..	8	7	..	..	4	3
Series 9	.	.	10	..	15	16	20	21	..	26	30	31
Series 14	.	.	32	..	..	..	28	29	26	27	..	25

Pressey, who requires the crossing-out of the number that breaks the rule of the series, gives these three samples:

(*a*)	.	.	2	4	6	8	9	10	12
(*b*)	.	.	7	6	5	1	4	3	2
(*c*)	.	.	1	2	4	8	16	17	—

Among the logical tests might be included those which demand the checking of numbers, names, or diagrams. I give the first two and the last two pairs of numbers in test 5 of the army Beta examination—the examination for illiterates:

650	650
041	044
41181900726	41181900726
6543920817	6543920871

The examinee is asked to mark with a cross those numbers that are identical. The National Intelligence Identity Test includes names and diagrams

as well as numbers. The instructions and three examples are here given :

If the two things in a pair are the same, write S on the dotted line between them. If they are different, write D.

307561 307561
Bateson, N. P. . . . Batterson, N. P.

 . . . . . 

Although in a way all intelligence tests are—or ought to be—tests of common sense, there is one kind of test that appropriates the label. The third of the Alpha Tests is known as a "common-sense test." The examinee is directed to mark with a cross, as in the sample, the best answer to a given question :

SAMPLE :

Why do we use stoves ? Because
- ☐ They look well.
- ☒ They keep us warm.
- ☐ They are black.

Then follow sixteen items of the same kind, the first of which is :

1. Cats are useful animals, because
- ☐ They catch mice.
- ☐ They are gentle.
- ☐ They are afraid of dogs.

And the last is :

16. Why is it colder near the poles than near the equator ? Because
- ☐ The poles are always farther from the sun.
- ☐ The sunshine falls obliquely at the poles.
- ☐ There is more ice at the poles.

We thus see that the tests begin with a very common form of common sense, but pass towards the

close into that organised form of common sense which we call science.

The old scholastic machinery of genus, species, differentia, proprium, and accidens appear in a new guise in the American tests. One of the most popular of these is the Logical Selection test. The instructions are: In each sentence draw a line under the *two* words that tell what the thing *always* has. In other words, the subject has to distinguish between the differentiæ and propria on the one hand and the accidentia on the other. Here are a few examples from the Terman Tests and from the National Intelligence Tests:

A bird always has
bones egg beak nest song.
Paper always has
edges envelopes printing surface watermark.
Fire always has
ashes danger flame heat wood.
Compromise always involves
adjustment agreement friendship respect satisfaction.

Another common classification test requires the removal of an intruder from among a group of things of the same class. Instructions: In each line cross out the word that does not belong there.

Examples: Frank James John Sarah William.
hard rough smooth soft sweet.

Otis presents the classification test in another way. Thus:

Directions: Find the way in which the first three things on a line are alike. Then look at the five other things on the same line and draw a line under the one that is *most* like the first three.

Sample: Hat, collar, glove . . . hand cane head shoe house.

I append two items :

President, captain, general . . ship, army, king, republic, soldier.
Generous, kind, honest . . . strong, selfish, wise, loyal, rich.

Roback in his Mentality Tests for Superior Adults gives two logical tests, one of which he calls an Abstraction Test and the other a Subsumption Test.

This is the Abstraction Test :

Directions : Place in front of each line the name of the general class to which *all* the words enumerated on the line can be referred.

Illustrations : A musical composition —song, oratorio, symphony, opera, cantata.

A few items :

: —Piano, compass, lancet, violin, pen.
:—¶ ÷ & ! √.
:—Asylum, hospital, college, library, jail.

The Subsumption Test is introduced thus :

Directions : Use each word in the following list as the first term of a series which you are to construct so that each successive term will be more inclusive or general than the last, until you have reached the most inclusive possible. For example, if the word given were *precinct*, the next term might be *ward* (as including *precinct*) ; the next *city* (as including *ward*) ; the next *county* (as including *city*), and so on—*state*, *country*, *world*, *universe*. Similarly the word *window-pane* might form the following series : pane, window, room, house, block, street, city, etc. Don't stop if there is any possible way of continuing the series. If you find yourself in a blind alley, take another path.

Then follows a list of initial words such as second (of time), leaf (of a tree), leaf (of a book), tone, family, etc.

Here it is evidently the relationship of whole and part that the author has in mind, not the subsumption of a species under its proximate genus.

Sometimes it is time sequence that is the relationship which the subject is asked to indicate. The following examples of time sequence are taken from the Dearborn Group Tests:

Directions: Number the following words to show their proper order. Put the numbers directly under the words as in the following examples:

A, dinner supper breakfast.
2 3 1

B, fruit flower seed plant.
4 3 1 2

I quote a few of the items:

Dress, silkworm, cocoon, satin.
Physician, football game, recovery, bandage, broken leg, tackle.
Armistice, battles, declaration of war, peace, recruiting, victory.

The relation of cause and effect is made the basis of 40 questions in the Miller Mental Ability Test.

Directions: You are to think of the first word in each group as a *cause*, and draw a line under *one noun* that *may be an effect* of it. (This is the author's wording, not mine.)

Samples:

FIRE (hot, house, damage, melt).
WATER (cold, flood, thirst, transparent).
ELECTRICITY (wire, snow, light, hot).
WORK (tired, honest, toil, wages).

I will close this chapter with two complete tests (A and B) in arithmetical reasoning, which I have myself used and found very serviceable. The tests, for each of which eight minutes are allowed, are approximately equal in difficulty. The age norms for each test are as follows:

Age	9	10	11	12	13	14
Score	4	5	6	7	8	9

These norms can easily be remembered from the formula :

$$N = \text{Age} - 5.$$

In using these norms, and indeed in using any of the norms in this book, it must be borne in mind that unless a complete age group is examined the results obtained cannot be validly compared with the norms. If, for instance, the top class in an elementary school be given one of these arithmetic papers, the average score should come out much higher than nine ; for the best brains in the school gravitate towards the top class, and about half the children of fourteen years of age are enrolled in the lower classes. If this important fact is ignored my norms will be found much too low.

ARITHMETIC

A

Eight Minutes.

Work these sums in your head. Make no mark on this paper.

1. A boy was given 10 apples, 3 of which were rotten and had to be thrown away. After eating 4 of the good ones, how many were left ?

2. Fred plays marbles, starting with 15. He loses 8 and then wins 6. How many has he then ?

3. Jane has a box of 24 chocolates which she is allowed to eat at the rate of 3 a day. How many days will they last ?

4. A man takes 20 minutes to walk from his house to the railway station. His son also takes 20 minutes. How long will it take them if they both walk together.

5. Find the number a quarter of which is 5.

6. If I pay 6*s.* for 3 lb. of butter, what shall I have to pay for 2 lb. ?

7. After spending half my money and then half the remainder I had 2*d.* left. How much had I at first ?

8. What is the least number that must be added to 53 to make it exactly divisible by 7 ?

9. How many eggs at three for 5*d.* can I buy for half a crown ?

10. If it takes three men to paint the inside of a house in two days, how many men would be needed to do it in half a day ?

11. A man works 2 days then rests 1, then works 2 days and rests 1, and so on. For every day he works he earns 10*s.* How much will he earn from Monday morning to Saturday night ?

12. What is the distance right round the edge of an oblong table 5 ft. long and 4 ft. broad ?

13. If I share a shilling between two boys so that one has 2*d.* more than the other, how much does the more fortunate boy get ?

14. Seven posts 3½ ft. apart are fixed in a row. What is the distance from the first post to the last ?

15. John and Henry start walking to meet one another from places 10½ miles apart. If John walks at the rate of 3 miles an hour, and Henry at the rate of 4 miles an hour, how long will it be before they meet ?

16. If a man's salary is increased by 10 per cent. and then reduced by 10 per cent., state whether he loses or gains, and by what percentage.

ARITHMETIC

B

Eight Minutes.

Work these sums in your head. Make no mark on this paper.

1. John had 12 marbles, 4 of which he gave to a friend and 3 were lost through a hole in his pocket. How many had he left?

2. Mary had a shilling, spent 4*d.* and had 6*d.* more given to her. How much had she then?

3. Among how many boys can I share half a crown so that each gets 3*d.*?

4. If it takes a stone 3 seconds to fall from the top of a tower to the ground, how long will it take two stones that start together?

5. Find the number half of which is 14.

6. If I have to pay 7*s.* for 2 lb. of coffee, how much will 3 lb. cost me?

7. After spending half my money and then half the remainder I have 3*d.* left. How much had I at first?

8. What is the least number that must be added to 40 to make it exactly divisible by 6?

9. How many oranges at 4 for 3*d.* can I buy for 2*s.*?

10. If it takes a man 3 days to paint the inside of a house, how many men would be needed to do it in half a day?

11. A man works a day, then rests a day, then works a day, then rests a day, and so on. For each day he works he earns 15*s.* How much will he earn from Monday to Friday night?

12. How far is it right round the edge of an oblong flower bed which is 5 ft. long and 4 ft. wide?

13. If I share 6*d.* between 2 boys so that one gets a 1*d.* more than the other, how much does the more fortunate boy get?

14. If telegraph poles stand in a straight row 50 yards apart, what is the distance from the first to the eighth?

15. A man walking at the rate of 4 miles an hour, pursues another man who has had an hour's start and walks at the rate of 3 miles an hour. How long will it take the pursuer to catch the other man?

16. If a man's salary is reduced by 10 per cent. and then increased by 10 per cent., state whether he loses or gains, and by what percentage.

CHAPTER VII

ABSURDITIES

As I have myself made extensive use of absurdities as group tests, both in the free form and in the fool-proof form, I will deal with this type of test in some detail.

Binet seems to have been the first to employ absurd statements as a test of intelligence. He used the following five absurdities, which he allotted to the eleventh year in the 1908 scale, and to the tenth in the revised scale of 1911 :

(1) One day a man fell on his head off a bicycle and was instantly killed. He was taken to the hospital and they say he will never get better.

(2) I have three brothers—Paul, Ernest, and myself.

(3) Once the body of a poor girl was found in a wood, cut into eighteen pieces. They say that she killed herself.

(4) Yesterday there was a railway accident; but the newspaper says it is not a serious one as only forty-eight people were killed.

(5) A man once said, " If I should ever grow desperate and kill myself I shall not choose a Friday to do it on, for Friday is an unlucky day and would bring me bad luck."

In order to pass the test the child must detect three out of the five absurdities.

By way of criticism it may be remarked that four out of the five deal with murder or sudden death; that three depend on the same basal fact (the fact that a dead man is dead and can neither act nor

suffer); that they are not equally easy; that they are not arranged in precise order of difficulty; and that there is some doubt as to the year to which they should be allocated. Burt considers No. 3 harder than No. 4, and No. 5 so hard that it is more suitable for children of eleven or twelve. Terman keeps the test in the tenth year, but Goddard and Kuhlmann place it in the eleventh year. Terman, however, makes two substitutions. Omitting No. 2 because in some languages the expression is idiomatically correct, and No. 5 because it is too hard, he substitutes the following:

(1) A man said: "I know a road from my house to the city which is downhill all the way to the city and downhill all the way back home."

(2) An engineer said that the more cars he had in his train the faster he could go.

To compensate for the easier items Terman requires four to be correct instead of three.

Of the other absurdities which have from time to time been suggested as alternatives to Binet's the best are these:

(1) This morning I met a smart young man. He was walking down the street with his hands in his pockets, and twirling a brand-new walking stick. (Whipple.)

(2) When there is a collision the last carriage of a train is usually damaged most. So the guard thinks it would be best if the last carriage was always taken off before the train starts. (Yerkes.)

(3) An Irishman called one day at the post office and asked if there was a letter waiting for him. "What is your name?" asked the postmaster. "Sure," said the man, "you will find my name on the envelope."

(4) A boy wrote in his composition: 'Soap smells nice but tastes horrid. It tastes worst of all when you get it in your eye."

The value of this type of test is beyond question.

Those who have extensively used it agree with Terman, who says: "The detection of absurdities is one of the most ingenious and serviceable tests of the entire scale. It is little influenced by schooling, and it comes nearer than any other to being a test of that species of mother-wit which we call common sense." And again: "It is an invaluable test for the higher grades of mental deficiency."

Valuable as these particular tests are for detecting the dull-witted and for marking a certain stage of mental development, they are of little use in differentiating ordinary adults. Terman observes that by twelve years of age critical ability is so far developed that the test is nearly always passed. This may be so if it is given in the form prescribed by Binet; but if we alter the mode of administering the test we get quite different results. Binet's absurdity tests are oral and individual; the child is told beforehand that each sentence contains something that is silly, and he is asked to discover it. But if, on the other hand, the child is left in doubt whether the statement is sensible or silly; if he is told that it may be either the one or the other, and is asked to say which and give his reasons, and especially if he is asked to write those reasons down, the difficulty of the task is greatly increased. Binet's plan gives two possibilities—to find the absurdity or not to find it. The other plan gives three—to judge the sentence sensible, or to judge it absurd without being able to explain the absurdity, or to judge it absurd and to explain the absurdity.

The difference is exemplified by a test which I have frequently used: "Captain Cook made three voyages round the world. In one of these voyages

he was killed by savages. Which voyage was it—the first, the second, or the third ? " Although not precisely conforming to the type described above, this is clearly an absurdity test admitting of three possible answers ; and if, when the third voyage is given a reason is demanded, the number of possibilities is increased to four. It is difficult to conceive that a lad of thirteen years of age with his wits about him could fail to give the right answer and the right explanation. Yet as a matter of fact about 40 per cent. of them do fail ; at least they do so when the test is written and not oral. Few give the wrong voyage (although about 5 per cent. do even this), but those who select the third voyage cannot always show that they decide on logical grounds.

I therefore set about devising a series of absurdity tests which would be free from the main objections urged against the Binet absurdities, which would vary in difficulty and in the type of incongruity, and which could suitably be applied to older pupils.

The tests I used in an investigation comprise 34 absurdities mixed with a few quasi-absurdities—statements which at first blush seem absurd but are not really so. I give below the 34 real absurdities arranged in increasing order of difficulty. They are derived from various sources. Some I have remembered, some (not many) I have invented, and some have been suggested to me by friends.

Instructions.—On the top of the paper write the name of your school, your name, and your age in years and months.

Some of the following sayings are sensible and some are foolish, and you must say which of them are sensible and which are foolish. Do not write out the saying, but simply put its number on your paper and write after it the word " sensible " or " foolish." It

you write the word "sensible" you need say nothing more; but if you write the word "foolish" you must give a reason for your opinion.

Section A

(1) A soldier writing home to his mother said: "I am writing this letter with a sword in one hand and a pistol in the other."

(2) It is said that a certain town in Greece contains two relics of St. Paul; one his skull when he was a boy, and the other his skull when he was a man.

(3) An old gentleman complained that he could no longer walk round the park as he used to: he could now only go half-way round and back again.

(4) In the year 1915 many more women got married than men.

(5) A hunter who had used up all his ammunition was chased by a bear. A bright idea struck him: he would climb a tree. When he got to the top he remembered that the bear could also climb a tree; but he got out of his difficulty by pulling the tree up after him.

(6) Light comes from the sun; feathers are light; therefore feathers come from the sun.

Section B

(7) A boy who wanted to go to a cinema, but had no money, thought it would be a good plan to walk in backwards, for the man at the door would think he was going out, and would not ask for his ticket.

(8) A teacher said to his boys: "To-morrow we hold a scripture examination, attendance at which is voluntary; so if any boy is absent he had better look out."

(9) A soldier in the march complained that every man in the regiment was out of step except himself.

(10) A sailor who was put to haul in a rope from the sea found it so long that he gave it up in disgust, saying that somebody must have cut the end off.

(11) The moon is more useful than the sun, for it gives us light in the night when we really need it, while the sun gives us light in the day when we don't need it.

(12) This is a sad and bitter world: we never strew flowers on a man's grave until he is dead.

(13) A stammerer who came from Birmingham to live in London

was asked why he stammered more in London. He replied: "B-b-bigger town."

(14) A man said to his shoemaker: "You blockhead; I told you to make one of the shoes larger than the other, and instead of that you have made one of them smaller than the other."

(15) A speaker at a war economy meeting said: "We must all economise. The man who generally buys four suits of clothes a year should be satisfied with three, the man who buys three should be satisfied with two, and so on."

Section C

(16) There is a tree in America which is so tall that it takes two men and a boy to see the top.

(17) A householder saw an advertisement: "Buy one of Simkin's stoves and save half your coal." He bought two in order to save all of it.

(18) A forgetful man once tied a knot in his handkerchief to remind him of something; but he suddenly remembered that the last time he did this he could not recall what it was he had to remember. So this time he tied two knots to make sure. The first knot was to remind him that he had to remember something; the second to remind him of what that something was.

(19) I am not conceited, for I don't think I am half as clever as I really am.

(20) A countryman came up to London during the war and saw on a hoarding the words: "Eat less bread: do it now." He immediately went into a tea-shop to do it now.

(21) "I don't like onions," said the boy, "and I'm glad I don't; for if I did I should be eating them all day, and I hate the beastly things."

(22) An Irishman was charged at the police-court with having stolen a pig. Five witnesses swore that they had seen him do it; but the prisoner said that he could prove his innocence by bringing 50 witnesses who would swear that they had not seen him steal the pig.

Section D

(23) "I have always noticed," remarked the old man, "that if I do not die in March I am all right for the rest of the year."

(24) Some years ago it was proposed to shift Sunday from the end of the week to the middle so as to divide the week into two parts.

(25) John James, who had married his widow's sister, used to say that if a man had a bad sister it was his misfortune, but if he had a bad wife it was his fault.

(26) You are thin, and I am thin; but he is thinner than both of us put together.

(27) One should never use long words when short ones will do just as well, for when the words are long the meaning is obscured by the unusualness of the terminology, and intelligibility is consequently endangered.

(28) The three men laughed; then stopped suddenly as the eyes of each met those of the others across the table.

(29) Watching pigs wallowing in the mire, he muttered: "No wonder they are called pigs."

Section E

(30) A man who bought a dog that had been advertised complained to the seller that the dog's legs were too short. The seller replied: "They are long enough to reach the ground, aren't they? what more do you want?"

(31) A showman advertised for a giant and a dwarf. A man of ordinary height presented himself and offered to fill both parts. He claimed to be the smallest giant in the world and the biggest dwarf.

(32) While standing near a clock tower just before the clock struck twelve two boys tried to find out which of them could hold his breath the longer. Neither of them won; for one was able to hold his breath from the first stroke of the clock to the sixth, and the other from the sixth to the twelfth.

(33) Every rule, even this one, has an exception.

(34) The horse obeys his master because his eyes magnify, so that his master seems to the horse to be much larger than the horse himself.

The plan I adopted in setting the test was to give each child a cyclo-styled copy of 38 statements comprising the 34 real absurdities interspersed with four spurious absurdities. The four spurious ones, arranged in increasing order of difficulty, were:

(1) A man may live many years on a crust—the crust of the earth.

(2) An old woman said that she was so old that in a few years' time she would be twice as old as she was forty years ago.

(3) I bought my little boy a watch which was said to give the right time twice a day; it turned out to be a dummy watch with fixed hands.

(4) The commonest word in the English language that is spelt the same way backwards and forwards is "a."

Any others of the same kind would have served just as well, for whatever answer was given it was not scored. Only the real absurdities were considered in estimating the results.

The time allowed was unlimited. The subjects were asked to hand in their papers as soon as they were finished.

The sectional classification of the absurdities is based on the percentage of adults who succeeded in detecting them. Each question in Section A was correctly answered by over 80 per cent. of adults, each in Section B by between 60 and 80 per cent., each in Section C by between 40 and 60 per cent., each in Section D by between 20 and 40 per cent., and each in Section E by less than 20 per cent. Only 6 per cent. dealt successfully with the last question.

The test was given to all children over eleven in two elementary schools, two central schools, three secondary schools, one day continuation school, one evening school, and one training college. Through the kind help of Dr. R. R. Rusk I was able to secure results from a few schools and a training college in Scotland. Altogether about 2,000 subjects were tested.

The papers were all marked by myself, or under my supervision, in accordance with the following key:

KEY

1. *Right:* "He had no hand free for writing." "Both hands occupied," etc.
   Wrong: "He couldn't write with a sword."
2. *Right:* "No man has two skulls." "A man has the same skull as he had as a boy," etc.
3. *Right:* "The distance is the same."
   Wrong: "He was a silly old man."
4. Some may contend that there is nothing inherently absurd in this, since a widow, widower, bigamist, or a divorced person may disturb the balance. Such an answer is to be considered correct.
5. Two points: (i) "The impossibility of pulling up the tree if he was on it," (ii) "The futility of doing it." If one point is given the test is passed.
   Right: "He could not pull up the tree if he was on it."
   "He had nowhere to put his foot except the tree."
   "He had no fulcrum."
   "If he pulled the tree up he would fall down with it."
   Wrong: "He could not pull the tree up."
6. The logical fallacy must be detected.
   Right: "The word 'light' is used in two senses."
   "When we say feathers are light we mean weight."
   Wrong: "Feathers do not come from the sun."
7. The answer must imply that the boy would still be moving inwards.
   Right: "The man at the door would see that the boy was going in, not out."
   Wrong: "The man at the door would stop him."
8. *Right:* "Voluntary means, if you like."
   Wrong: "The teacher should not punish a boy for being absent."
9. *Right:* "He was out of step himself."
10. *Right:* "By cutting the end off the rope is made shorter."
    "The rope still has an end."
    Wrong: "You can't cut the end off."

11. *Right:* "It is the sun that makes the day."
"Without the sun there would be no day."
Wrong: "The moon gets its light from the sun."
12. *Right:* "A man has no grave until he is dead."
13. *Right:* "The size of the town makes no difference."
14. *Right:* "You cannot have one larger without having the other shorter."
"Both statements mean the same."
15. *Right:* "The man who had one suit a year would have to be satisfied with none."
16. *Right:* "Two men and a boy cannot see farther than one man or one boy."
"We cannot add sights together."
Wrong: "Anybody could see the top."
17. *Right:* "Two stoves would burn more than one."
"To save all his coal he must burn none at all."
18. *Right:* "Two knots would not make him remember better than one."
"If the first knot would not remind him the second would not either."
Wrong: "Knots can't make you remember."
19. *Right:* "He contradicts himself."
"He first says that he is not conceited and then says that he is."
"He says he does not think himself clever and then says that he does."
Any answer will do that points out a contradiction, (*a*) between the two clauses or (*b*) between the two parts of the second clause.
20. *Right:* "He could not eat less by going into a tea-shop: he could only eat more."
"The advertisement asked you to eat less at meal times: the 'now' did not mean at this moment."
21. *Right:* "The boy assumes that he likes onions and that he does not like onions at the same time."
"He contradicts himself. He says, If I like them I should hate them."
22. *Right:* "The 50 witnesses were not real witnesses: they did not see anything."
"Only those who were present could give real evidence."

Any answer will do if it points out that negative evidence is in this case of no value.

23. *Right:* "He might have said the same of any other month in the year."
The fallacy consists in asserting that to be true of a particular month which is equally true of any other month.

24. *Right:* "There would still be a week from Sunday to Sunday."
"It would not split the week."
Wrong: "You can't shift Sunday."
"Sunday is the beginning of the week, not the end."

25. *Right:* "If a man has a widow he is dead."

26. *Right:* "Putting men together makes them fatter not thinner."

27. *Right:* "The writer does not follow his own advice."
"He tells us to use short words, and he himself uses long ones."

28. *Right:* "One man's eyes cannot meet two other men's eyes at once."

29. *Right:* "Pigs are not so called because they are dirty, but it is the other way about."
"Pigs is their proper name, their dirtiness has nothing to do with it."
Any answer will do if it suggests that the name is prior.

30. *Right:* "Every dog's legs reach the ground" (implying that the buyer's objection to this particular dog is not met).
"Legs half an inch long would reach the ground, but they would be of no use."
"The buyer meant that the legs were not long enough to be useful or beautiful."
"What the buyer meant was that the legs were not long enough to keep the body some distance from the ground."

31. *Right:* "A giant must be above the ordinary height, and a dwarf below the ordinary height."
"Smallness is a bad quality in a giant, and largeness in a dwarf."

32. *Right:* "The second boy won, for he held his breath for six intervals, while the first held it for five."

33. *Right:* "Self-contradictory."
"This itself is a rule and the exception to it is that there is a rule which has no exception. Therefore if the rule is true it is also false."

34. *Right:* "If his eyes magnify they magnify everything in the same proportion."
"If he sees a man big he also sees another horse big, and sees his own body big."

This series of tests was devised and used for a specific purpose—the purpose of discovering the age at which intelligence ceases to grow. I started the investigation with a belief that it continued to grow right up to the prime of life; I ended it with a belief that there was but little increase after 15 or 16 years of age. The evidence for this belief, and the various interpretations that can be put upon the facts, will be discussed in a later chapter. It will suffice here to give the average score obtained from the whole number of children, youths, and maidens whom I tested.

Averages:	13·1	14·4	15·1	17·4	18·5	18·9	18·9
Age:	11	12	13	14	15	16	17

As two out of the four elementary schools tested were central schools, the norms for the ages 11 to 14 are higher than would be obtained if all the children of those ages in large areas had been tested. The same is true of the higher ages, for these norms were mainly secured from secondary schools where we may fairly assume that only the brightest pupils remain till they are 17.

Those who wish to repeat my experiments with the same tests can use either the whole series or any selection that would present the same difficulty as the whole series. As the questions are arranged in

order of difficulty, this selection may be made by taking alternate items, or every third item, or the middle item of each section, or two items from each section equally removed from the middle of that section, or any other arrangement that contains a balanced number of easy and difficult items. Having made a selection, however, it would be well to adhere to it for the whole inquiry; for the series is not so carefully standardised that the items rise in equal steps; selections made in accordance with the above instructions will be approximately, not exactly, of equal difficulty. The objection to using a small number of items is that a few bright subjects may get them all right. This is not likely to occur if the full series be used. None of my own subjects got more than 32 marks out of the total of 34. The absurdities should be mixed with a few pseudo-absurdities, either those given above or any others of a similar kind; but the pseudo-absurdities should be ignored in marking the papers.

The percentage of absurdities detected and explained gives the grade of intelligence. The normal adult seems to get about 50 per cent. of the marks.

The same tests put into fool-proof form are included in the set of intelligence tests which appear in Chapter XIV.

The aspect or mode of intelligence which absurdity tests call into play seems to be the power to construct coherent wholes. It is that synthetic power which Bosanquet[1] regards as essential for inference. "Ultimately," he says, "the condition of inference is always a system . . . a group of

[1] *The Essentials of Logic*, 140.

relations of properties or things so held together by a common nature that you can judge from some of them what the others must be." Some indeed require the system to be apprehended as a spatial and concrete thing before they are satisfied. Lord Kelvin[1] did not at first accept Clerk Maxwell's electromagnetic theory of light, because he could not make a model of it. An absurdity test contains an incongruity—something that does not fit; and the failure to fit becomes apparent only when all the pertinent elements are brought together in the mind to form a whole. In these tests all the elements of knowledge requisite for a solution are just as familiar to the youngest testee as they are to the oldest. The oldest is robbed of his advantage of a larger store of knowledge, so that any difference in the response is due to a difference in synthetic power—provided the mental attitude is the same. If the attitude of the testee is one of suspicion; if he looks upon the test as a "catch" and not a *bona fide* question; if he is so wary that he avoids the answer that appears most obviously right, it is clear that the results cannot be taken at their face value. When this frame of mind is suspected steps should be taken to dispel it before the test is given.

[1] *Life of Lord Kelvin*, by Silvanus P. Thompson, ii. 835-6.

CHAPTER VIII

GENERAL KNOWLEDGE

SINCE it is customary to draw a sharp distinction between intelligence and knowledge, a test of pure knowledge is the last thing we should expect to find in an intelligence examination. Yet here it stands as one of the eight Alpha Tests, one of the ten National Intelligence Tests, and one of the six Haggerty Tests. And the knowledge tested is not always universal knowledge: it is often knowledge peculiar to one country, one people, and one period. Binet's tests, it is true, test knowledge; but then it is knowledge common to the civilised world. The scale would suffer no essential loss if it were translated into Chinese and used to test the children in the schools of far Cathay. But the Alpha Information Test is moored to New York and Chicago. The first of the 40 items—America was discovered by Drake Hudson Columbus Balboa—refers to a matter of universal knowledge, and so in fact do several other items; but let the English reader try his hand at these:

2. Pinochle is played with rackets cards pins dice.

6. Food products are made by Smith & Wesson Swift & Co. W. L. Douglas B. T. Rabbitt.

9. Marguerite Clark is known as a suffragist singer movie actress writer.

17. Nabisco is a patent medicine disinfectant food-product tooth paste.

18. Velvet Joe appears in advertisements of tooth-powder dry goods tobacco soap.

The National Intelligence questions are much more general, especially in the anglicised version. Some of them remind us of Magnall's Questions and carry us back to the days when *The Child's Guide to Knowledge* was an indispensable school book. The questions, however, are less exotic, and are certainly much easier, for it is often unnecessary to remember the right answer: it is sufficient to recognise it, or even to infer it. One of the questions, for example, runs thus:

The Plymouth Rock is a kind of cattle fowl sandstone horse.

If the examinee were inclined to believe that the Plymouth Rock is a sweetmeat or a religion, the alternatives offered would at once dispel the belief, and, if he had ever heard the term applied at all, could scarcely fail to strike the right association. A process of more complete elimination would enable him with very imperfect knowledge to answer another question in the same test:

Treasure Island tells about Black Dog Fagin Ivanhoe Oliver Twist.

He can give the right answer if he has never read Stevenson at all, but is merely familiar with Dickens and Scott. The opportunity thus afforded of making the most of what knowledge one happens to possess is, we may presume, the only excuse for regarding these tests as intelligence tests at all; unless indeed we frankly accept the view that a test of knowledge is also a test of intelligence.

CHAPTER IX

TESTS FOR ILLITERATES

A LARGE number of soldiers in the American Army were so illiterate that they were unable to take the Alpha examination. For them was devised an alternative series of tests, known as the Beta examination. As many of these illiterates were foreign immigrants, the examiner had to explain the nature of the task to be performed, by gesture and by demonstration on a blackboard.

The first test was similar in kind to Porteus's

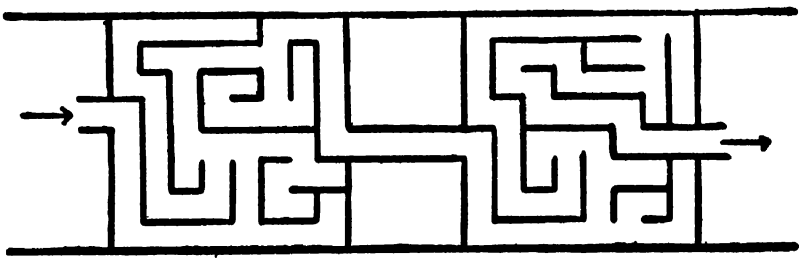


FIG. 1.—MAZE TEST.

well-known mazes. The examinee had to trace with a pencil the shortest way through the maze. Of the five mazes set, the fourth is given in Fig. 1.

The second test involved the counting of a pile of little cubes represented by a drawing in perspective on the examination paper. The twelfth and thirteenth of the sixteen items are represented in

Fig. 2. The number of cubes had to be inserted in the squares beneath.

The third test required the completion of 12 patterns made up of noughts and crosses. It is

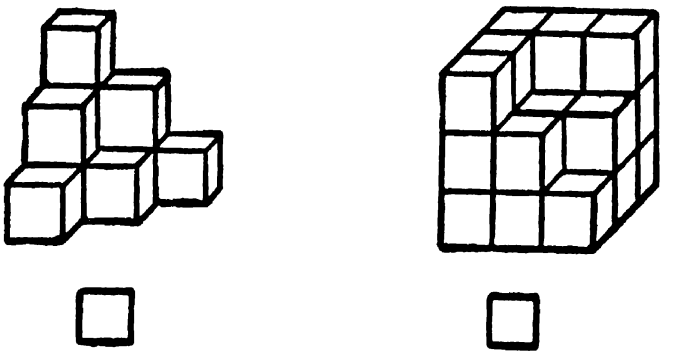


FIG. 2.—CUBE ANALYSIS TEST.

illustrated in Fig. 3, which represents the eighth and ninth items.

The fourth was a substitution test. The subject was given at the top of his paper the key reproduced

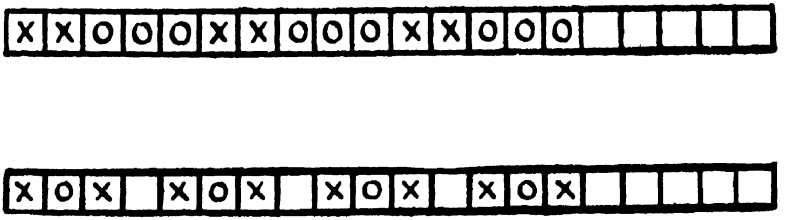


FIG. 3.—X-O SERIES.

in Fig. 4 and was required to place in the squares under the digits the corresponding symbols. The first of six lines of digits is here given.

The fifth test required the checking of fifty pairs of numbers—numbers which sometimes ran into

eleven figures—by placing a X against those that were exactly the same (see p. 36).

The sixth test was based on *Binet's* missing-features test. The page in the *Beta booklet* con-

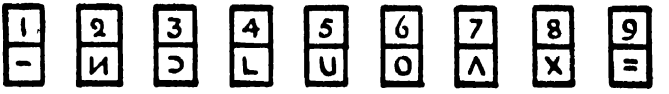


FIG. 4.—DIGIT-SYMBOL TEST.

tained 20 little pictures in each of which something essential was missing. This the examinee was required to supply. The illustration in Fig. 5 gives four of these pictures.

The seventh and last test consisted of ten squares,

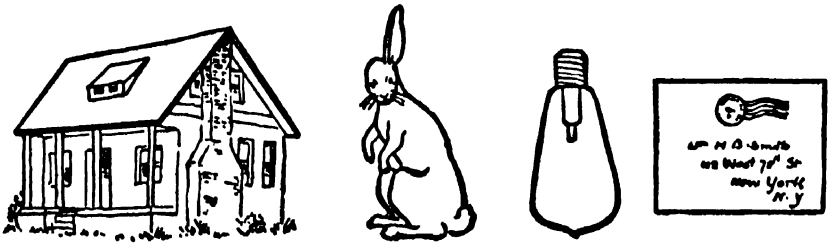


FIG. 5.—PICTORIAL COMPLETION TEST.

each of which could be dissected into, or built up from, other geometrical figures placed by its side. Lines had to be drawn in the square to show how it could be constructed from the given elements. I illustrate the last two examples in Fig. 6.

The times allowed for the seven tests were 2, 2½, 1¾, 2, 3, 3, and 2½ minutes respectively.

Among the illiterates we must include children under 10 years of age, for although most children of 8 and 9 can read and write they cannot do so with the necessary facility and the necessary uniformity. Group tests, especially if there is a narrow time limit, presuppose that the pupils take approximately the same time to read the questions, and, where there is writing to do, to write the answers, so that the

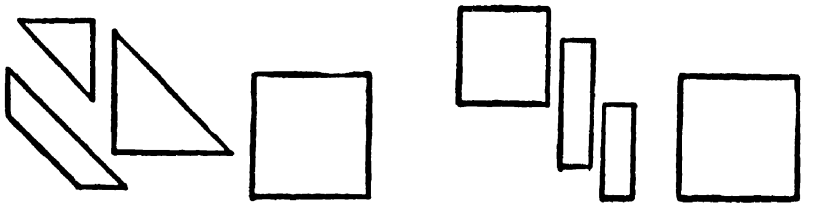


FIG. 6.—GEOMETRICAL CONSTRUCTION TEST.

differences in the achievements are wholly due to differences in the speed and accuracy of thought. But children under 10 vary so much in speed of reading and writing that this pre-supposition is never in point of fact realised. If therefore they are given tests of the pencil and paper type, those tests must be concerned with pictures, diagrams, and figures rather than with printed words. It follows, too, that the instructions must be oral. Tests of this kind are somewhat rare. The best known are the Otis Primary Examination and the Haggerty Intelligence Examination Delta I. There are a few others, however, which seem to be equally good, such as the Myers Mental Measure; the Pressey Cross-out Tests, Primer Scale; the Detroit First-grade Intelligence Test; and the Kingsbury Primary Group Intelligence Scale.

The instructions type of test occurs in a variety of forms in all these scales. The child is asked to do something to a picture or group of pictures—to insert a missing part; to delete an irrelevant, or correct an absurd, item; or simply to identify and mark a certain object, or to draw a line from one part of the picture to another. As a good example of this kind of test I give in Fig. 7 the second row of little pictures in the first test of the Otis Primary Examination.

FIG. 7.—INSTRUCTIONS TEST (OTIS).

The instructions given by Otis for this row of pictures are as follows:

"(1) Now look at this row of pictures and draw a circle round the doll. (Pause 5 seconds.)

(2) Next find the picture of something that can run, and draw a line under it. (Pause 5 seconds.)

(3) Next find the picture that is between the doll and the candle and make a little cross under it. (Pause 5 seconds.)

(4) Next find the picture of something that gives light and can be picked up. Make a round dot under it. (Pause 5 seconds.)

(5) Next draw a line from the teddy bear's ear to the rabbit's ear that will go under the sun. (Pause 5 seconds.)

(6) Next find the picture of a child's plaything that has large ears, and put a little circle under it. (Pause 10 seconds.)"

Haggerty complicates his later instructions by inserting the usual hypothetical clauses. The tenth and last item in exercise 2 consists of five squares in a row, for which the following instructions are given:

"Put your finger on 10. Look at me. When I say 'Go,' if a pig is larger than an elephant put a cross in the last square; but if not put a cross in the first square. Ready! Go!" (10 seconds allowed.)

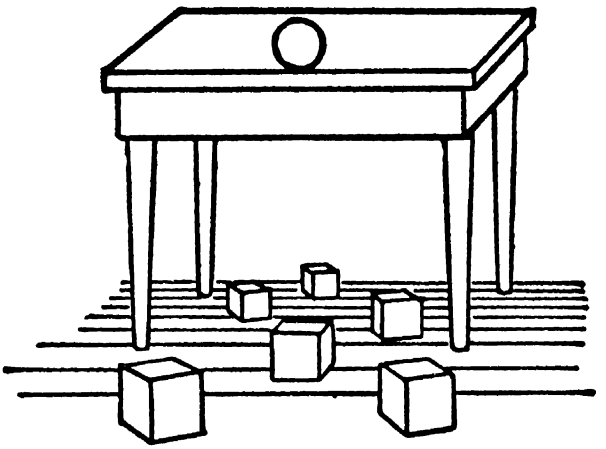


FIG. 8.—INSTRUCTIONS TEST (KINGSBURY).

Kingsbury gives several ingenious instructions tests, one of which is illustrated in Fig. 8.

This is what the examiner has to say about this drawing: "Here is a table with a ball on the edge of it, and some blocks on the floor below. If the ball should roll off the table, which block would it strike? Mark a cross on the block it will strike." (Allow 5 seconds.)

Besides the instructions test there are other varieties of the printed verbal group test which appear in pictorial form. The opposites test, for instance, is given by Kingsbury in the form repre-

sented in Fig. 9. The pupil is required to draw a circle round the picture of the object that is more different from the one in the square than any of the others.

A picture-checking test is sometimes used in the

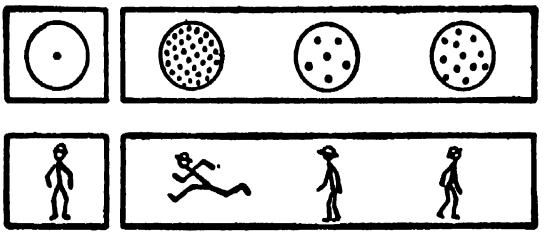


FIG. 9.—PICTORIAL OPPOSITES (KINGSBURY).

same way as a number-checking test. Haggerty has a test which consists of little pictures arranged in pairs right down the page. If the members of the pair are the same, the child has to place an S between them, if they are different he has to place a D.

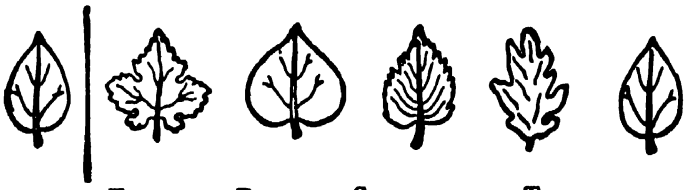


FIG. 10.—DETROIT SIMILARITIES TEST.

Another form of similarities test which appears in the Detroit First-grade Intelligence Test is illustrated in Fig. 10. The pupil is directed to mark the first leaf and then mark the other leaf that looks just like it.

Otis's method of presenting the similarities test is illustrated in Fig. 11. The pupil is directed to look at the first three pictures and see how they are

alike; then he has to select from among the other five pictures the one that is most like the first three, and put a cross under it.

Similarities pass into classification. Miss Engel,

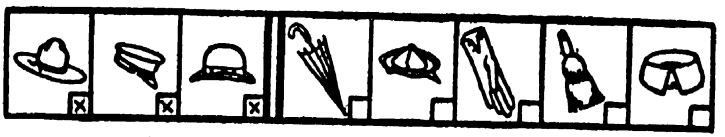


FIG. 11.—OTIS'S SIMILARITIES TEST.

who devised the Detroit scale, presents a series of tasks like the one illustrated in Fig. 12. The pupil is asked to mark three things made of wood.

The Myers Mental Measure, although it consists

FIG. 12.—DETROIT CLASSIFICATION TEST.

entirely of pictures, is claimed by its author to be applicable to all ages. It certainly has a much wider range than the other picture scales and runs into higher degrees of difficulty. About half the tests

FIG. 13.—MYERS' CLASSIFICATION TEST.

are classification tests, of which Fig. 13 gives a specimen.

The examinee is told that there are four things, and only four things, in the row that are alike in some way, and he is asked to draw lines under them. The keyword for this row is "bipeds."

Haggerty uses a felicitous form of the digit association test, the key for which is given in Fig. 14. The test requires the insertion of the right digits under a long and mixed series of the pictures.

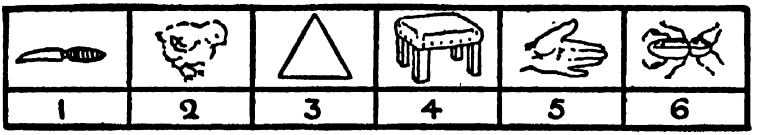


FIG. 14.—HAGGERTY'S ASSOCIATION TEST.

It will be observed that the relationship between the number and the pictures is not arbitrary, but such that an intelligent child would immediately

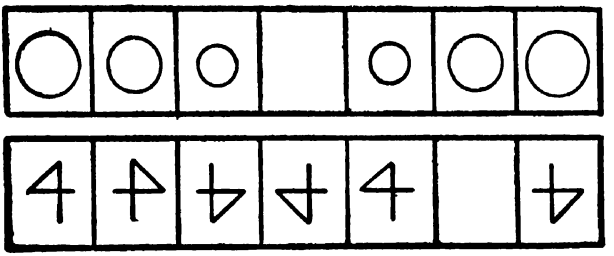


FIG. 15.—KINGSBURY'S COMPLETION TEST.

discern and thus be saved the time and trouble of repeated reference to the key.

Fig. 15 illustrates two items of Kingsbury's com-

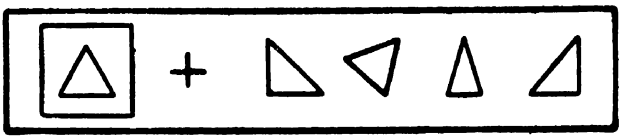


FIG. 16.—KINGSBURY'S FORM TEST.

pletion test. The pupil is asked to fill in the blanks so as to complete the series.

Kingsbury also presents an interesting variety of the geometrical form test. The pupil is required

to find among the four blocks to the right the one which when fitted into the block to the left will make the latter a solid square. Fig. 16 exemplifies this test.

The Primer Scale of the Indiana Mental Survey

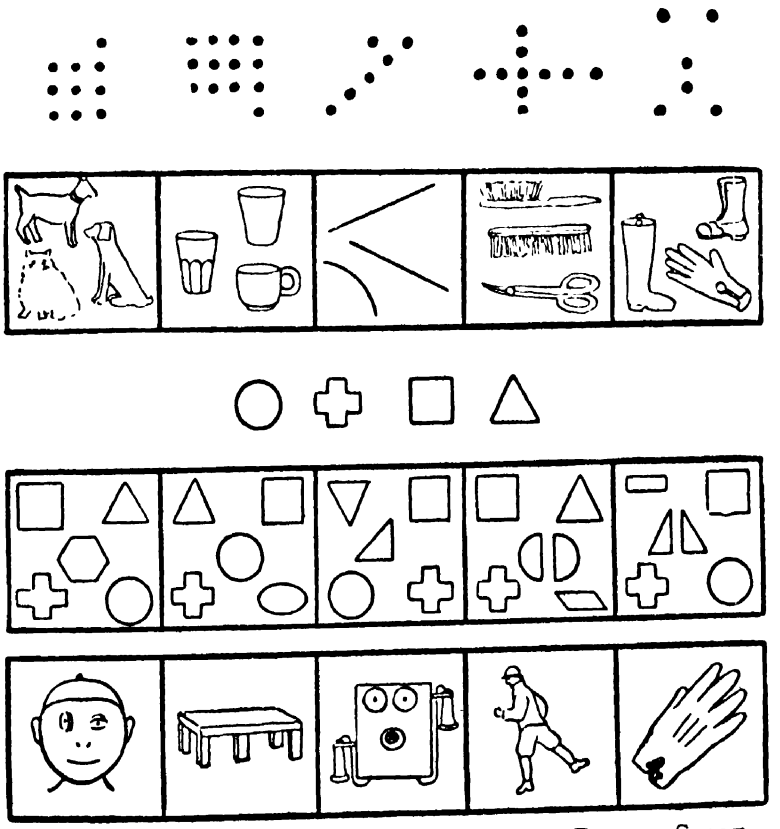


FIG. 17.—THE PRESSEY CROSS-OUT TESTS—PRIMER SCALE.

Tests consists of pictorial analogues of the Pressey Cross-out Tests. Mrs. Pressey thus describes these tests in the *Journal of Educational Psychology* for September 1919.

"The scale consists of four tests, each of twenty-five items—a total of one hundred items in all.

In addition there are five examples for each test. The first row of items for each test is reproduced (in Fig. 17). In the first test, the children are asked to cross out the 'extra' dot. In the second, they are told to cross out, in each square, the thing that is different from the other two things in the square. In the third, they are to cross out, in each square, the block that will be left over after all the other blocks have been fitted into the four forms at the top of the page. And, in the fourth, they are to cross out in each picture the part that is wrong." Three minutes are allowed for each of the four tests.

CHAPTER X

GROUP TESTING IN ENGLAND

ENGLAND is not yet alive to the full import and significance of mental tests. Her psychologists have not seriously attacked the problem of testing intelligence by simple and expeditious methods over a large field. Such group tests as have been used here in past years have not been designed as general scales of measurement, but as means to the solution of special educational problems. Mr. W. H. Winch has derived many excellent group tests, but they are *ad hoc* tests—tests for one particular purpose. The nearest approach to a general scale is his set of reasoning tests. Mr. Cyril Burt, who has devoted himself more exclusively to the measuring of intelligence, has given us a number of valuable group tests which are likely to be widely used. The Bradford Education Committee has recently used some of them for the selection of scholarship children. And the Northumberland Education Committee has followed suit and asked Professor Godfrey Thomson to devise for the same purpose the set of tests which will be described in the next chapter.

English psychologists have never taken very kindly to the time tests; nor are they quite reconciled to the fool-proof test. They suspect the time test of favouring a superficial smartness, and they sur-

mise that in making a test fool-proof we deprive it of much of its value. Since, however, the relative merits of the various types of testing have never been experimentally determined, the questions at issue are still regarded as open, and American methods are being cautiously and vigilantly tried.

A noteworthy stage in the progress of mental testing is marked by the conversion of the Civil Service Commissioners. In the Civil Service competitive examinations for women clerks in October 1920 there appeared for the first time a section headed Intelligence Tests. The question paper contained 14 tests, one on each page. Exactly two minutes were allowed for each test. Every two minutes a bell was rung and the candidate had to turn over a new page and tackle a new test. Marks were allowed for every item done right and deducted for every one done wrong. So guessing was not encouraged; nor was it better to have tried and failed than never to have tried at all. As the general structure and pattern of this paper has been followed in subsequent Civil Service examinations, it is worth while to describe it in detail.

The first test is an instructions test of which I quote two items out of six:

4. In the following list of words draw circles round three so that two mean things belonging to the same class whilst the third means that class:—

food mat ton salt Sunday cake

6. Draw a continuous line passing under the first word in the following list, through the second, and over the third, and then repeat the operation when

the words in the list are counted from the other end and the instructions under and over exchange places :

head cap tree box

Test II was similar. It contained six instructions, of which I quote the fourth :

Write on the left of the second line of this sentence the first letter that in this sentence goes next before the letter which in the alphabet goes next after it.

Tests III and IV are free analogies—analogies as originally used by Mr. Burt, with no suggestion as to the missing term. The next two tests (V and VI) are controlled analogies : the candidate has to choose one of four possibilities. The seventh and eighth tests are completion or missing-word tests. The ninth and tenth are the same in fool-proof form. The ninth, for instance, consists of a continuous piece of prose which begins thus :

foolish
Inability to take the tide at the turn is wrong.
fatal

purpose
Many a man owes his whole battle in life to the
success

see
faculty of being able to keep the golden moment
win

and catch it before it is past.

It continues thus for about 200 words. The candidate is told that where three words, or groups of words, are given one above the other, he has to underline the one that makes the best sense.

Test XI begins with a prose passage of about 50

or 60 words, which is followed by five simple questions on its meaning. It is, in fact, a silent reading test. Test XII is similar.

Though all the tests are admirably devised, it is not till we come to Test XIII that we find a marked element of novelty. The other tests are all original in respect of content, but not in respect of form. Test XIII is original in form as well. It begins thus:

For each question, see what one, two, or three (not more) consecutive words of the passage make the best answer. Under each of these words write the number of the question. The first is done, as an example.

Just as wildness is found in some youths but not in all, so too avarice sets its brand unmistakably on a small number of old men, but not on the rest.

I I I

Every age has vices to which it is specially liable, but none to which it is bound by fatal necessity.

1. How many old men are here found fault with?

2. What fault is said to occur in some young men?

Four more questions follow. Test XIV is the same.

There is manifest throughout the whole paper the influence of America, no less in the form of the test than in the mode of procedure. The time limit is imposed throughout, and all but four of the tests are fool-proof.

One of the questions set at a later examination came perilously near a hoax: Describe the wedding-dress of Queen Elizabeth.

Similar to the Northumberland Tests in general structure and in mode of procedure is the Simplex

Group Intelligence Scale by Mr. C. A. Richardson. It has, however, a larger variety of tests and takes about twice as long to work.

Under the enterprising leadership of Mr. T. G. Tibbey, the Education Research Committee of the London Head Teachers' Association have experimented with group tests of intelligence and have pronounced them sound and practicable.

Group tests with important bearings on the theory of intelligence have been carried out by the psychological staff of University College, London, under the direction of Professor Spearman.

CHAPTER XI

THE NORTHUMBERLAND MENTAL TESTS

THROUGH the courtesy of Professor Godfrey Thomson (the author) and Messrs. Harrap & Co. (the publishers) I am able to present in full the Northumberland Mental Tests, which seem to me to be the most promising of recent attempts to improve upon scholarship examinations.[1] The primary object of the tests was "to discover gifted children worthy of free secondary education among elementary schools of the county of Northumberland which had not this year sent in any candidates for the orthodox examinations in English and mathematics on which such free scholarships are usually awarded." Twenty scholarships were presented on the results of the inquiry. The tests, however, were given not only to the 414 scholarship candidates, but also, for purposes of standardization and comparison, to nearly 3,000 other children. One of the most interesting facts brought to light was that the most intelligent children in the county are to be found not in the large towns,

[1] For a complete account of the experiment and its results, see the *British Journal of Psychology* (General Section), vol. xii, Part 3, December 1921, pp. 201–22.

nor yet in the small country towns and villages, but in the remote country places—especially among the Cheviots. Yet at the usual scholarship examination these children cannot successfully compete with children who attend efficient urban schools, but who are considerably inferior in natural ability.

Although the tests are presented, as the American tests are, in pamphlets which constitute both a question paper and an answer paper, there is a wide difference in the details of presentation. The American paper has to be worked through piecemeal, each test (that is, each group of similar questions) having its own time limit, and the examiner, with stop-watch in hand, giving instructions in accordance with a fixed and fairly elaborate programme. The Northumberland examination is, in comparison, simplicity itself. All the supervisor has to do is to see that the names and ages are properly given, and that the children have exactly one hour to work the paper. So while the American tests are "fool-proof" at one end—the marking end—the Northumberland tests are "fool-proof" at both ends; it is almost impossible to make a mistake in administering them, almost impossible to make a mistake in marking them.

Professor Thomson has, no doubt, acted wisely in rejecting the time limit for individual tests; but he has in so doing renounced one marked advantage of the time limit—the advantage that the candidate is compelled to distribute his time reasonably over the whole paper. This difficulty has, however, been foreseen and forestalled. The six tests have been so divided and arranged that it is

almost impossible for any one of them to be neglected.

To prevent nervousness, and a sense of strangeness, at the examination a preliminary practice test was given the day before. All the necessary instructions were printed on the papers themselves: none had to be given by the supervisors.

The practice test is given on page 79, and the examination proper on the pages that follow. Some of the tests will be familiar to the reader; but there are two—the Hindustani Test and the Extra Number Test—which were invented by Professor Thomson; and there are two others—the Middle Word Test and the Schema Test—which, though based on the work of Professor Stern of Hamburg, are new in the form of presentation.

Professor Thomson's account of his scheme of marking is as follows:

The rule for marking the tests was in general that only the *exact* answer gained points. This requires interpretation in several cases, however, as given on pages 88 and 89.

There are 60 items in the tests in all, and for simplicity the marking was made quite straightforward and the score was the number of items correct.

The rules prescribed will appear arbitrary in some instances, but they are based on an examination of a large number of cases and appear to be more justifiable than other alternatives. It must be remembered that all the tests are in part directions tests; disobeying directions should therefore lose points even if the answer is intelligible.

THE NORTHUMBERLAND MENTAL TESTS

You have ten minutes for this test

PRACTICE TEST

Look at these words:

Roof : house :: hat : head.

You see that a roof does to a house what a hat does to a head. Look at these:

Sheep : wool :: cat : ?

The fourth word is not given, but you know it must be fur. Sheep have wool and cats have fur. Look at these:

Swimming : water :: flying : gun, air, cork.

Here three words are printed in the fourth place, and you have to pick the right one and underline it. (It is air.) Now try to underline the proper word in each of the following lines. (Underline the words heavily and plainly so that they are easily seen.)

Snow : white :: grass : army, coal, green.

Eat : bread :: drink : iron, water, stones.

Sailor : navy :: soldier : fight, gun, army.

July : month :: Friday : day, week, year.

Success : failure :: joy : pleasure, work, sadness.

Flour : miller :: soot : sweep, black, corn.

North : south :: left : east, right, behind.

Frame : picture :: lake : window, island, photograph.

Knowledge : ignorance :: light : darkness, wisdom, red.

Known : unknown :: present : gratitude, weeping, future.

Sorrow : misfortune :: joy : grief, happiness, success.

Add : subtract :: multiply : increase, divide, add.

Potato : vegetable :: veal : calf, fruit, meat.

Character : reputation :: truth : brave, opinion, lie.

The tests you will have will of course not be quite the same as this, and they will last an hour instead of ten minutes. But like this test they can all be done by following the instructions exactly, and everybody will be able to do some of them.

THE NORTHUMBERLAND MENTAL TESTS

The Supervisor should check age and birthday.

Name_____

Age *last* birthday_____

Birthday_____

School_____

Standard_____

To-day's date_____

Wait for the signal before you turn over.

There are some easy questions inside and you have to answer them as quickly and carefully as you can. Begin at the beginning and go straight through. If any one is too hard for you, miss it out. You have an hour to do them in. If there is no clock in the room you will be told the time every quarter of an hour. Stay till the end.

Ask no questions at all.

Test A1

In the tests given below you have to *cross out the extra word* in each line. For example look at this first line:

wood cork stone boat bladder

The extra word there is stone, because all the others float on water. Look at this second set:

chair table stool desk roof

The extra word there is roof because all the others are articles of furniture. Now try to cross out the extra word in each of the remaining lines:

grapes oranges wool apple banana
rifle plough sword pistol lance

Test A2

You have to give the number that comes next in each of the following lines of numbers. The first three are answered for you to show what is meant:

1 2 3 4 5 6 (7)

Here the number that you have to write in the brackets is 7 because the first six numbers go up by steps of one at a time. The series below also have steps of different sizes but they do not always increase by adding the same number each time, but in other ways as well:

11 10 9 8 7 6 (5)

Here the numbers come down.

1 2 4 8 16 32 (64)

Here each number is twice as big as the one in front.

Now try the others:

3 6 9 12 15 18 ( )
2 4 6 8 10 12 ( )
1 3 5 7 9 11 ( )
1 2 3 4 3 2 ( )

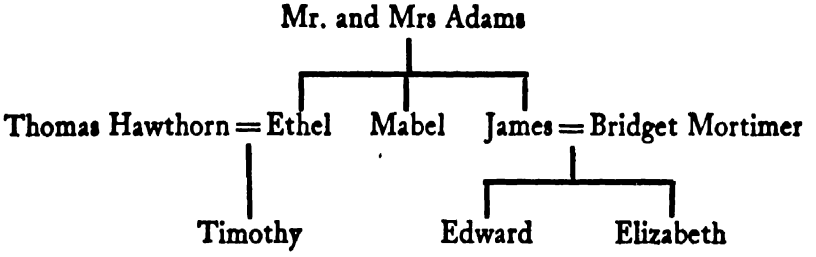
Test A3

Read the following paragraph carefully and compare it with the diagram. The diagram gives the same facts as the paragraph.

PARAGRAPH

Mr and Mrs Adams had three children, (1) Ethel (the eldest), (2) Mabel and (3) James (the youngest). Ethel married Mr Thomas Hawthorn, and their son was named Timothy. James married Miss Bridget Mortimer, and they had two children (1) Edward and (2) Elizabeth.

DIAGRAM



Now answer these questions (the first two are answered for you to show what is wanted):

	Answers
What relation is Timothy to Mabel?	Nephew
How many aunts has Edward?	Two
What is Timothy's surname?	
What is Elizabeth's surname?	
Who is Timothy's uncle?	

Test A4

In your mind (*without writing them down*) you have to arrange the five words in each line below in their proper order, and then underline the middle word of this order. For example, consider

minute second year hour week

In their proper order these words would run thus : second, minute, hour, week, year. The middle word is hour and therefore hour is underlined above. Similarly shilling is the middle word of this next set:

sixpence penny florin sovereign shilling

Now try to do the next tests in this way.

elephant	sheep	mouse	cow	puppy
forty	thirty	fifty	ten	twenty

Test A5

You have to cross out the "extra" number on each of the following lines. For example, in

6 2 8 7 4 10

you cross out 7 because it is the only odd number, all the rest agree in being even numbers.

In this next case

4 7 8 6 19 3

you cross out 19 because it is the only number in double figures. Again in

14 3 15 9 6 12

you cross out 14 because all the rest are divisible by 3. Try to find and cross out plainly the "extra" number in each of the following lines :

18	16	4	8	20	12
5	9	3	4	1	7
75	62	20	10	15	25
17	49	4	24	13	18

Test A6

The sentences below are in a foreign language, and their meanings are given in English. In each English sentence a word is underlined, and you have to underline the word which corresponds to it in the foreign sentence. You can do this by comparing the sentences with each other. For example, look at these:

1. Kuchh malai	some cream.
2. Kuchh puri leoge	will you take some cake ?
3. Misri leoge	will you take sugar ?

By comparing 1 and 2 you see that kuchh must mean some because it occurs in both sentences. Underline it in sentence 1 before you go any further. By comparing 2 and 3 you see that leoge means will you take. Underline it in 3. Then the only word you do not know in 2 is puri which must mean cake, so underline it in 2. You have now underlined in each foreign sentence the word corresponding to the underlined part of the English sentence.

Notice that the foreign words are not always in the same order as the English words.

Underline your words plainly.

You have not to write anything, only to underline the proper words. Now try these:

Ek piyala chae	A cup of tea.
Yih chae bahut achchhi hai	This is very good tea.
Chae bilkull taiyar hai	Tea is quite ready.
Kab taiyar karoge ?	When shall you make ready ?
Main bahut pyasa hun	I am very thirsty.
Bahut achchhi hai	It is very good.
Yih mera rumal nahin hai	This is not my handkerchief.

(You may turn back to read examples if you wish to do so.)

Test B1

Cross out plainly the "extra" word in each of the following lines:

charity	kindness	benevolence	revenge	love
square	circular	oblong	hexagonal	triangular
needle	tack	nail	knife	pin
coal	bread	coke	wood	paper
bran	wool	cotton	hemp	jute
hair	feathers	wool	grass	fur

Test B2

Give the number that comes next in each of the following lines:

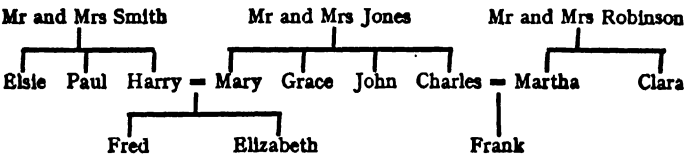
1	3	9	27	81	243	(	)
17	15	13	11	9	7	(	)
1	2	4	7	11	16	(	)
96	48	24	12	6	3	(	)
7	5	5	7	5	5	(	)
9	8	7	6	7	8	(	)

Test B3

Cross out the "extra" number in each of the following lines:

26	3	7	31	13	17
18	22	30	24	6	12
81	27	11	1	9	3
8	2	16	32	6	4
7	28	21	35	14	27

Test B4



From the above diagram answer the following questions:

	Answers
What is Frank's surname?	
What relation is Elizabeth to John?	
What relation is Elizabeth to Frank?	
What relation is Elizabeth to old Mr Jones?	
How many aunts has Frank?	
How many uncles has Fred?	
Who are Frank's grandparents?	

Test B5

PART 1

In your mind (*without writing them down*) you have to arrange the five words in each line below in the proper order and then underline the middle word of this order:

paragraph	book	chapter	sentence	word
house	street	room	town	county
orange	red-currant	grape	plum	melon
general	sergeant	captain	private	corporal

PART 2

In the remaining line two such sets of five words are mixed up together. You have to separate them mentally, arrange them in order, and underline the middle word of each. *Do not write them down*, just underline the proper two words.

leaf, leafvein, twig, sowing, reaping, thrashing, branch, ploughing, tree, baking.

Test B6

In each foreign sentence underline the word which corresponds to the underlined word in the English sentence.

Underline your words plainly. You have not to write anything, only to underline the proper words.

Mausam badalta hai	It is changeable weather.
Achchha khub mausam hai	It is good weather.
Sirf das baje hai	It is only ten o'clock.
Das gaz napo	Measure ten yards.
Ham kahan jaen ?	Where shall we go ?
Ham chha baje khama khawenge	We shall dine at six o'clock.
Yih rah kahan jati hai ?	Where does this road lead to ?

Look over the tests again, until time is up

The particular instructions are:

*Test A*1. Correct answers *wool*, *plough*, 2 points. No point if two words in a line are underlined, or if the words are written down, or if all the other words are crossed out and the proper one left. Underlining or ringing the word is accepted as crossing out. Alterations are allowed if they are perfectly clear. *These principles apply* mutatis mutandis *in the other tests.*

*Test A*2. Correct answers 21, 14, 13, 1. 4 points. The point is awarded also if the series is further continued correctly.

*Test A*3. Correct answers *Hawthorn* (or Timothy Hawthorn, or Mr Hawthorn, but *not* Thomas Hawthorn); *Adams* (or Elizabeth Adams); *James* (or James Adams, but *not* James Mortimer, or Mr Adams, or J. B. Mortimer). 3 points. Spelling mistakes are ignored if the meaning can be grasped.

*Test A*4. Correct answers *sheep*, *thirty*. 2 points.

*Test A*5. Correct answers 18, 4, 62, 4. 4 points (4 is the only number not in double figures).

*Test A*6. The language is a simplified Hindustani. Correct answers *chae*, *hai*, *bilkull*, *taiyar*, *bahut*, *achchhi*, *yih*. 7 points. No points are deducted if words in the English are unnecessarily underlined (such candidates seldom score more than a chance point in the seven). Marks are neither given nor deducted for the explanatory sentences (kuchh, etc.).

*Test B*1. Revenge, circular, knife, bread, bran, grass. 6 points.

*Test B*2. 729, 5, 22, 1½, 7, 9. 6 points. No point for 1 instead of 1½.

*Test B*3. 26, 22, 11, 6, 27. 5 points.

*Test B*4. Jones
niece (*not* uncle)
cousin (or cousins, *not* half-cousin)
granddaughter (or grandchild, *not* grandfather)
three
three
Mr and Mrs Jones and Mr and Mrs Robinson. 8 points, *i.e.*, 2 for the grandparents of Frank, thus:

one grandparent only	no point.
any 2 or 3 grandparents	1 „
4 grandparents	2 „
too many grandparents	no „

Robinsons = two grandparents.

Jones and Robinsons = 4 grandparents.
Jones and Robinson = 2 ,,
Jones (alone) = no point.

Test B5. Paragraph, street, plum, sergeant, twig, reaping. 6 points.

Test B6. Badalta, mausam, sirf, das, ham, baje, kahan. 7 points.

The norms of performance are as follows:

9½	10½	11½	12½	13½ years.
9	15	24	33	42 points

CHAPTER XII

A PICTURE TEST

As a test for children about 6 years of age I have made a series of drawings which have been printed on both sides of a sheet of paper. They are reproduced below. In giving the tests to a class each child is provided with a lead pencil and one of the picture sheets. Either the child or the teacher should fill in the blanks for name, age, etc. Then the class is instructed as follows (the directions to the teacher are put in brackets) :

" Look at the first picture. It is a kind of fork. Draw it as exactly as you can in the empty place by its side. [Indicate by holding up one of the sheets and pointing. If all have understood what you mean and have drawn the fork in the right place, proceed as follows :]

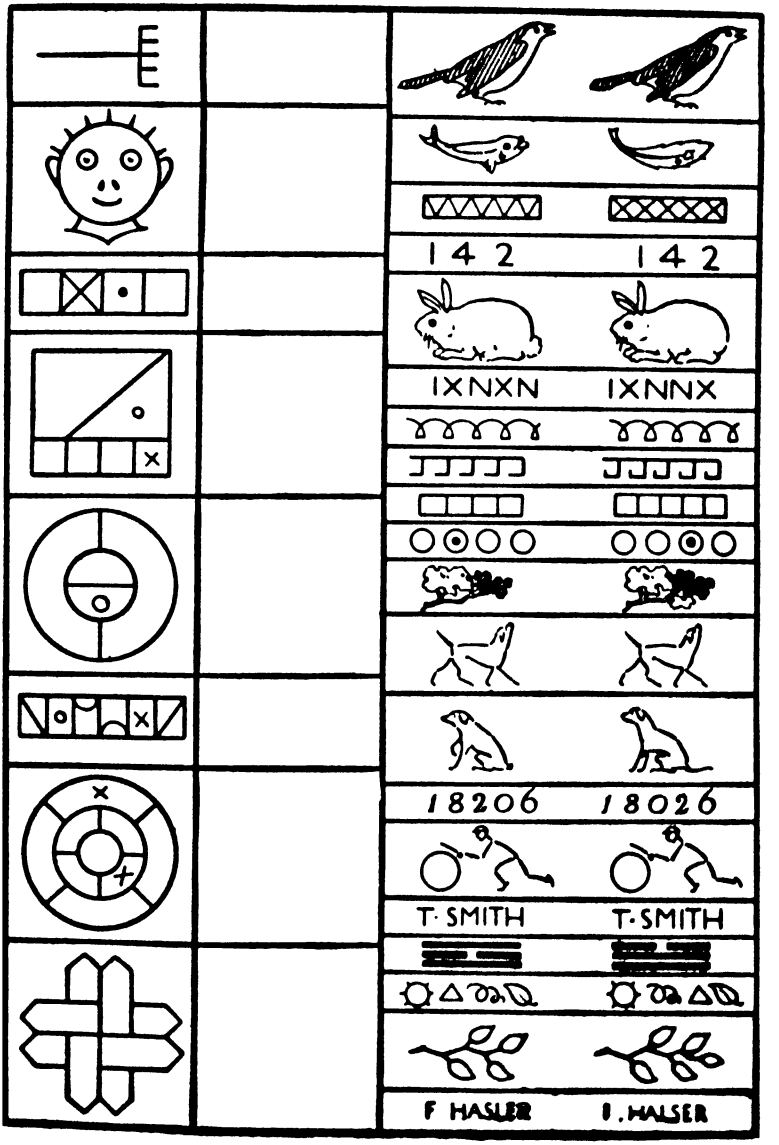
" Now look at the picture below it—the picture of a man's head. Draw it as exactly as you can.

" Then do the same with all the other drawings right down the page. [Point to the pictures and empty spaces. There is no time limit for these tests. Don't hurry the pupils; don't urge them to go slowly. Let them go their own pace; unless of course they are wasting time.]

" Now look at the pictures on the other half of the page. Some of these pairs of drawings and figures

Picture Test

School Name Sex Age



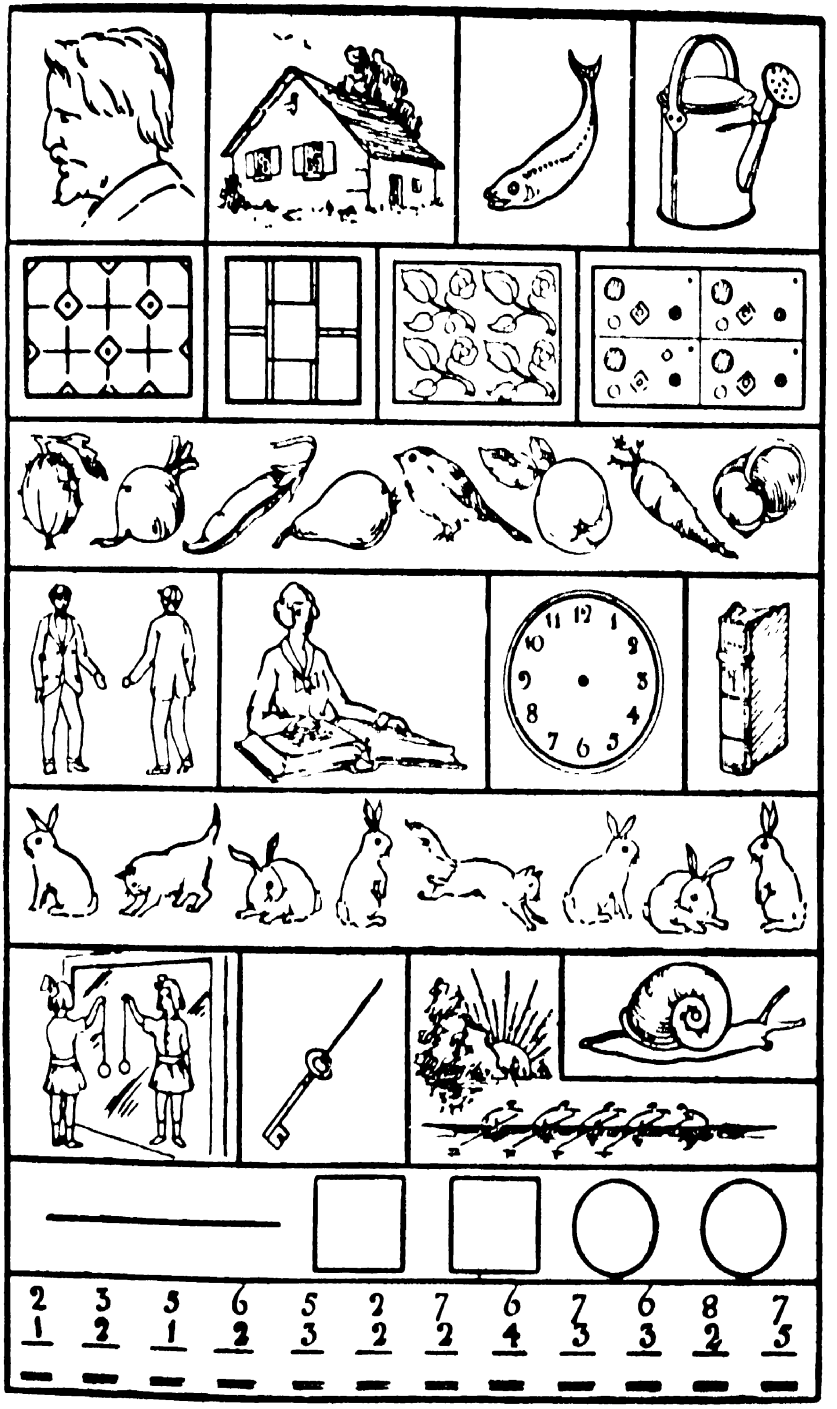
and words are exactly the same, and some are a little different—are nearly the same but not quite. When the two drawings are quite the same join them with a line; when they are different in any way put a cross between them to show that they are wrong. [Illustrate what is meant by making drawings of your own on the blackboard. When you think the pupils understand, say:] Now do it. [When they have finished, say:]

"Now turn your papers and look at the other side with a man's head as the first picture. In each of those pictures on the top line there is something missing. Put it in. [Draw some illustrations of your own on the blackboard, such as a man without a nose, or a kettle without a spout, and show clearly what the children have to do.]

"Now look at the next row. In each of these four pictures there is something too much, something that ought not to be there because it spoils the pattern. Put a cross on the part that is wrong. Begin at the second if you like, because it is the easiest. [Illustrate as before on the blackboard.]

[You must be careful with the next test. In the previous tests you may repeat your instructions until you are sure the children know what they are required to do. Here the instructions must be given *twice only*; and there must be no blackboard illustrations.]

"I am going to tell you to do something to the third row of pictures. You must listen very carefully, for I am going to say it twice only. Then you must do what I have said. Now listen: Draw a little line under each of the vegetables and put a cross on each of the fruit. Draw a little line under


2 3 5 6 5 2 7 6 7 6 8 7
1 2 1 2 3 2 2 4 3 3 2 5

each of the vegetables and put a cross on each of the fruit. [Pause till it is done.]

"Now listen again. I will tell you twice what to do. Draw a line from the beak of the bird to the turnip passing above the pear and below the pea-pod. [Repeat. Pause.] Now listen again. I will tell you twice what to do. Draw a line from the pear to the carrot passing below the bird and above the apple. [Repeat. Pause.]

"Look at the next row. Here are two men coming to meet each other. Put a stick in the right hand of each. [Pause.] The next picture shows you a lady reading a book. There are two pages open. Put a cross on the page she reads first. [Pause.]

"The next is a picture of a clock with no hands. Put in the hands showing half-past three. [Pause.]

"Look at the picture of the book. Put a little cross above the last page of the book. [Pause.]

"In the next row there are rabbits and cats. I want you to make the number of rabbits the same as the number of cats. If there are too many cats, cross out the extra cats; if there are too many rabbits, cross out the extra rabbits. Cross out with a thick stroke. [Pause.]

"Look at the picture of the little girl looking into the looking-glass. If you can see anything wrong, cross it out and draw it as it ought to be. [Pause for some time.]

"The next picture shows you a key swinging on a nail. Draw it as it will look when it swings round so as to come on the dotted line. [Pause.]

"Here are some men rowing in a boat. Put a cross at the front of the boat to show which way the boat is going. [Pause.]

"Above the boat is a snail with a shell on its back. If you see anything wrong, cross it out and draw it as it ought to be.

[Show on the blackboard how to cut a line into two parts by putting a little stroke in the middle. Then say :]

"Look at the line in the next row. Cut it into three parts, all the same size. [Pause.]

"Look at the first square. Divide it into four parts, all the same size and the same shape. [Pause.]

"Now do the same to the other square, but do it in a different way. [Pause.]

"Now divide the circle into four parts, all the same size and the same shape. [Pause.]

"Divide the last circle into *three* parts, all the same size and the same shape."

[Work this subtraction sum on the board with the class .—, and tell them that they will be given three minutes to work the 12 subtraction sums in the last row. Stop them as soon as the three minutes have expired.]

$$\begin{array}{r} 5 \\ 2 \\ \hline 3 \end{array}$$

DIRECTIONS FOR MARKING

The copied pictures are to be marked right if there is clear evidence that the child has observed accurately and has noted the number and relation of the parts. In the first drawing, if he has five prongs to the fork he has passed the test. In the second there must be exactly seven hairs on the head and no feature must be omitted. Bad drawing does

not matter so long as the various items are there in their right place and their right number. Very young children tend to multiply the items. They will give a dozen prongs to the fork and abundant hair to the man. In these drawings one mark is to be given for each of the first six, two marks for the seventh, and three for the eighth.

The score for the next twenty items is the number right minus the number wrong. For instance, if 16 are attempted (enough time should be given for all to be attempted) and 9 are right, the score is $9 - 7 = 2$.

On the next page the missing parts in the first row are the ear, the chimney, the fins (one fin suffices) and the handle behind. Each scores one mark.

In the next row the first pattern carries two marks and the rest one each.

Eight marks are given for the next row; one for each of the vegetables correctly indicated, one for each of the fruit correctly indicated, and one for each of the other two orders correctly carried out.

Two marks are given if the stick is placed in the right hand of each man; if of one only, no marks are allowed. The score is two or nothing. The next (open book) carries one mark, the next (clock) carries three, and the next (closed book) carries one.

Two marks are allowed if four rabbits are crossed out in the next row.

The mirror test is difficult. If the reflected arm with the ball is crossed out and drawn on the other side of the body, four marks are scored. The key carries three marks, the boat one, and the snail four. The snail's head should be the other side; or the shell turned round.

Each item in the next row is allowed one mark only, except the last, which is allowed three.

Each correct sum scores two.

The total number of marks amounts to 100.

The marker is recommended to keep by his side a spare copy of the test, with a bold figure in red placed on each picture which carries more than one mark.

I have not yet applied this test to a sufficient number of children to secure trustworthy norms.

Here are tentative norms:

Age	6	7	8	9
Norms	48	60	72	84

CHAPTER XIII

THE COLUMBIAN MENTAL TESTS

THE set of tests that follows is an attempt on my part to construct, on American principles (hence its name), a group test that should be applicable to English children. It is a survival—the final product of many trials and many changes. Behind it lie discarded tests and rejected methods. I first experimented with a set of time tests; but judging from the results it seemed as though they all tested the same sort of thing. Each echoed the verdict of the others. The time element was too prominent: the opportunity for reflection too scanty. I weeded out some of the tests, extended the time limit in others, removed it altogether from the "Best Answer" test, and added another untimed test of my own, which I labelled Common Sense. Six tests remained, four with time limits and two without.

Although the forms of five out of the six tests are old, the contents of all are new, in the sense that I made up the various items without referring to, or consciously recalling, tests already published. I cannot of course believe that none of the analogies, for instance, have ever been used before; and I know that most of the number series must have been used by many other examiners; but, speaking generally,

the items are new, and the compilation is a new compilation.

This group test, in spite of the name I have given it, differs from the American group tests in two important respects: it does not lay so much stress on the time factor, and it includes a larger number of tests that involve distinct elements of reasoning. The five minutes allowed for each of the first four tests is about double what the Americans would allow. In fact the examiner will find that some of the children will complete the tests before the five minutes have expired. But where this is most likely to happen (as in the analogies test and the number series test) the increasing difficulty of the items will of itself amply distribute the children without the aid of a brief time limit. Indeed the important differentiating factor is difficulty and not speed; yet speed is not left entirely out of account, for the first two tests are intended to gauge the child's capacity to think not only accurately but rapidly as well.

The main idea was to secure a balanced series—a series that would bring into play the various aspects of intelligence that bear upon school studies and to give the slow, tenacious thinker the same chance as the quick and volatile.

Any child who can read and write with fair facility is able to take the Columbian Tests. They are specially suitable for ordinary children between the ages of 10 and 14. If the children are older than this, or indeed if they are younger and are of high intelligence, they can be more fittingly examined by the Chelsea Tests, which will be described in the next chapter.

The Columbian Tests are set forth in a printed or

cyclostyled booklet, so arranged that only one of the first four tests is exposed at a time. This is essential in a series of time tests.

Each child is supplied with a lead pencil and a piece of paper about quarto size—a piece with at least twenty lines. This is his answer paper. He is asked to rule it as in the key given below, to fill in details of name and age, to place in the top row the numbers of the tests from 1 to 6, and in the first column the numbers of the items from 1 to 20. Then he is ready to begin.

The test books are then distributed with the warning that they are not to be opened until the order is given. Then the examiner says: "When you open your books (or papers) you will find Test 1 on the first leaf. It is called Obeying Orders, because it asks you to do 16 things. As you are allowed only five minutes to do the whole 16 you will have to do them quickly. They have to be done in the column headed Test 1, the first on the first line, the second on the second line, and so on. If you are asked to draw something, such as a square or a circle, draw it very rapidly like this [the examiner illustrates on the blackboard], bearing in mind that marks will not be given for neatness, but only for accuracy and smartness. Now turn over the first page and begin.

[At the end of five minutes the examiner says:] "Pencils down. Look at me. The next test consists of 16 sentences with the words all mixed up like this:

"Grow trees on apples. [The examiner writes this on the blackboard.]

"If you put the words in the right order so as to make a sensible sentence, what do they become? 'Apples grow on trees.' The last word of this sensible sentence is 'trees,' and this word is what you would have to put down on your answer papers. What you have to do is to find the last word of the sensible sentence. Now try to answer these:

"Always John best his does.
"On runs a rails tram.

[The examiner writes them on the blackboard, and assures himself that the class understands that the answers required are "best" and "rails."]

"Turn over to Test 2 and begin.

[5 minutes allowed.]

"Pencils down. See what I have written on the blackboard:

"*Table* is to *wood* as *window* is to . . .?
(frame, glass, putty, legs.)

"Choose from the words in brackets the one that goes with 'window' as 'wood' goes with 'table.' Glass is the right word, and is what you would have to put down on your answer paper.

"Now try this one:

"*Boots* is to *feet* as *gloves* is to . . .?
(leather, sleeves, hands, laces.)

"'Hands' is the word that should go down on your answer paper. Is there anyone who does not understand what to do? Very well; turn over to

Test 3 and begin. You must stop at the end of five minutes.

[5 minutes allowed.]

"Pencils down. Look at these numbers that I have written on the blackboard :

1 2 3 4 5 6

What would the next two numbers be? 7 and 8. Now find the next two numbers for each of these :

2	4	6	8	10	12
9	8	7	6	5	4
1	4	1	5	1	6

[If the children clearly see what is required they are told to start Test 4. Five minutes allowed.]

"Pencils down. See what I have written on the blackboard :

"A policeman wears uniform so as—

A. To be warmly clad.
B. To be easily recognised.
C. To look smart.
D. To frighten thieves.

"Which is the best reason why a policeman wears uniform? The one marked B. Therefore B should be written down on your answer paper.

"Now find the best reason here :

"People wear rubber soles on their boots because :

E. They are fashionable.
F. They make no noise.
G. They make the soles last longer.
H. Rubber is more porous than leather.

"The reason marked G is the best, therefore your answer should be G. Now answer in the same way the sixteen questions in Test 5. You will be given as much time as you like, so think carefully. When you have finished you can work the sixth and last test."

The papers should be marked in accordance with the given key. Each correct answer scores 1, making a total score of 100.

The following age norms have been secured:

Age	10	11	12	13	14	15
Norms	46	56	64	69	72	74

THE COLUMBIAN MENTAL TESTS.

TEST I—OBEYING ORDERS

Five Minutes

1. Print the first letter of the alphabet.
2. Make a cross and put a ring round it.
3. Draw a square and print the letter S in it.
4. Place three crosses in a row. Join the first with the last by a line which passes above the middle one.
5. Make a circle. In it put a W; outside on the right put a T, and on the left of the circle put a C.
6. Write the capitals X Y Z in a row. If 8 is less than 3, cross out Z; if not, draw a line under X.
7. If a square is rounder than a circle, draw a circle in a square; if not, draw a square in a circle.
8. Write the last letter of the word in this extraordinary sentence that has more than twelve letters.
9. Draw side by side a triangle, a circle, and a square. In the triangle put the letter T, in the circle put nothing, and in the square put any number that is the wrong answer to the question: twice four?
10. Write down the number of letters in the fourth word in this sentence unless a cow is bigger than a cat, in which case write the word "Rat."
11. If a dog has two noses and one eye, write the word "yes"; if he has two eyes and one nose, write the word "no."
12. Write the number that comes before the number that comes before fourteen.
13. Draw a square and a circle so that half the circle falls within the square. In the part of the circle that is outside the square write the number of feet in a yard.
14. If Liverpool is in Scotland write the word "Dublin"; but if it is in England, write the word "Scotland."
15. Draw a triangle unless there are more days in the week than there are weeks in a month, in which case draw a circle.
16. If this sentence contains more words of less than three letters than words of more than three letters, write the first letter of the last word; if it contains less, write the last letter of the first word.

TEST II—MIXED SENTENCES

Five Minutes

Make the sentence sensible and write down its last word:

1. Are round apples.
2. Grow bushes blackberries on.
3. An trunk a has elephant.
4. Little can many skip girls.
5. Pigs fat some very are.
6. All soldiers well good march.
7. Not do pudding like plum Frenchmen.
8. Dull make sleepy feel one books.
9. The head worn on are hats.
10. Carried sea the is coal across.
11. With is a bat played cricket.
12. Clear should be explanation an always.
13. Water to thirst us quench our enables.
14. Can penny little a bought for be very.
15. Generally those us who to are love we kind.
16. Some a pity do that is work it not like people.

TEST III—ANALOGIES

Five Minutes

Find the fourth word:

1. Leg *is to* knee *as* arm *is to* . . . ?
   (hand, wrist, elbow, sleeve)
2. Father *is to* son *as* mother *is to* . . . ?
   (sister, daughter, aunt, brother)
3. Far *is to* near *as* up *is to* . . . ?
   (above, below, under, down)
4. Snow *is to* white *as* ink *is to* . . . ?
   (black, pen, letter, writing)
5. Second *is to* two *as* third *is to* . . . ?
   (fourth, last, next, three)
6. Dog *is to* puppy *as* cat *is to* . . . ?
   (tail, pussy, kitten, purr)
7. Orange *is to* peel *as* nut *is to* . . . ?
   (pulp, kernel, shell, rind)
8. Fire *is to* smoke *as* water *is to* . . . ?
   (liquid, wet, ice, steam)
9. Sheep *is to* flock *as* bees *is to* . . . ?
   (honey, swarm, hive, sting)
10. Donkey *is to* bray *as* horse *is to* . . . ?
    (gallop, cart, neigh, reins)
11. Lamb *is to* gentleness *as* lion *is to* . . . ?
    (roar, prey, strength, ferocity)
12. Wax *is to* candle *as* oil *is to* . . . ?
    (can, lamp, motor-car, paraffin)
13. Singing *is to* speaking *as* poetry *is to* . . . ?
    (music, prose, recitation, drama)
14. Man *is to* manly *as* child *is to* . . . ?
    (womanish, childish, girlish, namby-pamby)
15. Manners *is to* morals *as* politeness *is to* . . . ?
    (politics, society, kindness, virtue)
16. House *is to* rent *as* capital *is to* . . . ?
    (labour, interest, country, money)

TEST IV—NUMBER SERIES

Five Minutes

Give the next two numbers in each row:

(1)	12	11	10	9	8	7
(2)	1	3	5	7	9	11
(3)	9	9	8	8	7	7
(4)	1	5	2	5	3	5
(5)	1	1	3	3	5	5
(6)	6	9	12	15	18	21
(7)	40	35	30	25	20	15
(8)	6	2	5	2	4	2
(9)	1	2	4	8	16	32
(10)	2	3	6	7	10	11
(11)	9	12	10	13	11	14
(12)	8	9	11	12	14	15
(13)	10	7	9	6	8	5
(14)	2	4	5	10	11	22
(15)	1	4	9	16	25	36
(16)	1	2	4	7	11	16

TEST V—BEST REASON

Write down the letter that stands before the best reason or the best answer.

1. When a little girl loses her doll she should—

   A. Cry till somebody finds it for her.
   B. Think where she is likely to have left it and look there.
   C. Search in her father's pockets.
   D. Ask her mother to buy her a new one.

2. If you are caught in a shower far from home and have no umbrella you should—

   E. Take shelter till the shower passes.
   F. Run all the way home.
   G. Ask a policeman to lend you an umbrella.
   H. Borrow money to buy a mackintosh.

3. Boys should not smoke because—

   K. It should be left for men to do.
   L. It is not healthy.
   M. Tobacco is expensive.
   N. Very few children ever smoke.

4. More birds die in the winter than in any other season because—

   O. They lose their feathers.
   P. They can't get enough water.
   Q. They can't get enough food.
   R. Boys throw stones at them.

5. Children like riding in roundabouts because—

   S. It makes them giddy.
   T. It is cheap.
   U. They like to be admired.
   V. The movement is pleasant and exciting.

6. Children read "comics" because—

   W. They contain pictures.
   X. They are cheap.
   Y. They make them laugh.
   Z. They are good for their morals.

7. They let children travel by rail at half price because—

A. They never injure the carriages.
B. They take up less room.
C. They are dear little things.
D. They don't earn money.

8. A man sneezes when he has a cold—

E. To make other people sorry for him.
F. To drive away the cold.
G. To make use of handkerchiefs.
H. Because he can't help it.

9. Shops are open late on Saturday night because—

K. People wish to buy things for Sunday.
L. People have a half-holiday on Saturday.
M. Shopkeepers like to make a lot of money on Saturday.
N. People go out on Saturday nights to see the shop windows.

10. People go to a cinema—

O. To learn about other countries.
P. To enjoy themselves.
Q. To meet their friends.
R. To eat oranges and nuts.

11. Carpets are useful because—

S. We can wipe our boots on them.
T. They collect and hide the dust.
U. They show that we are well off.
V. They are soft and warm.

12. A weathercock is placed on the church steeple to tell us—

W. What sort of weather it is.
X. Which is the north.
Y. When it is too stormy to go to church.
Z. Where the wind blows from.

13. Days are shorter in winter than in summer because—

A. Cold contracts.
B. The clouds hide the sun.
C. That part of the earth slopes more away from the sun.
D. The days are foggy and gloomy.

14. Poking the fire brightens it because—

E. It irritates the fire.
F. It lets in more air.
G. Iron is a good conductor of heat.
H. The poker gets hot.

15. We eat turkey on Christmas day because—

K. It goes well with plum pudding.
L. Turkeys are expensive.
M. It has become the custom.
N. Turkeys are intended to be eaten at Christmas.

16. The sea is salt because—

O. There are salt rocks at the bottom.
P. Salt fish swim about in it.
Q. People put salt there many years ago.
R. Rivers constantly carry a small amount of salt into it.

TEST VI—COMMON SENSE

John has a sister Jane, a brother Dick, and a cousin Tom. Answer the following six questions about these people:

1. Who is Dick's sister?
2. Who is Jane's cousin?
3. Who is Tom's female cousin?
4. Who is Dick's brother?
5. How many brothers has Jane?
6. How many cousins has Tom?

Five boys, Tom, Dick, Joe, Bill, and Sam, sat in a row. Dick sat at one end of the row and Joe at the other Bill sat next to Dick and Tom sat next to Joe. Now answer the following five questions about them:

7. How many boys sat between Dick and Joe?
8. Who sat in the very middle?
9. Who sat between Bill and Tom?
10. Who sat between Sam and Dick?
11. How many boys sat between Bill and Joe?

Now answer the rest of the questions:

12. Mary is bigger than Mabel, and Margaret is bigger than Mary. Who is the smallest of the three?

13. Sarah is as old now as Anne was two years ago. Which is the older?

14. How many great-grandmothers have you (never mind their being dead)?

15. Fred is as tall as William would be if William were twice as tall as he really is. Which is the taller, Fred or William?

16. If all bloaters are herrings, are all herrings bloaters?

17. Henry and Philip ate six apples between them. Henry ate twice as many as Philip. How many did Philip eat?

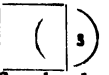
18. Five years ago Phyllis was two years older than Maud. How many years older than Maud is she now?

19. A cyclist rode a mile on an old-fashioned bicycle which had a big wheel in front and a little wheel behind. Which wheel went round the larger number of times—the big or the little?

20 Read the last question again. Which of the two wheels travelled the faster?

THE COLUMBIAN MENTAL TESTS

KEY

No.	Test i.	Test ii.	Test iii.	Test iv.	Test v.	Test vi.
1	A	round	elbow	6 5	B	Jane
2		bushes	daughter	13 15	E	Tom
3		trunk	down	6 6	L	Jane
4		skip	black	4 5	Q	John
5		fat	three	7 7	V	two
6		well	kitten	24 27	Y	three
7		pudding	shell	10 5	B	three
8	y	sleepy	steam	3 2	H	Sam
9		head	swarm	64 128	K	Sam
10	Rat	sea	neigh	14 15	P	Bill
11	no	bat	ferocity	12 15	V	two
12	12	clear	lamp	17 18	Z	Mabel
13		thirst	prose	7 4	C	Anne
14	Scotland	penny	childish	23 46	F	four
15		us	virtue	49 64	M	Fred
16	f	work	interest	22 29	R	no
17						two
18						two
19						little
20						neither or same

CHAPTER XIV

THE CHELSEA MENTAL TESTS

THE Chelsea Tests are so named because they were first used in that part of London which the old London School Board used to call the Chelsea Division. I prefer them to the Columbian Tests because they are more completely my own. A man naturally prefers his own children to his foster-children.

The set comprises four tests, each carrying 25 marks. The first is a cipher test—the only one with a time limit. The second is literary, and gauges the pupil's grasp of the meanings of words. The third consists of my absurdities test put into fool-proof form. The fourth, which I have called an orientation test, owes its origin to the researches of two friends, Dr. E. O. Lewis and Mr. Hugh Gordon. Dr. Lewis, who was for many years engaged in investigating mental deficiency, holds that the capacity to orientate is singularly lacking in the feeble-minded. They seem to possess, in very small measure, what may be called geometrical imagination. This interesting fact suggested the idea of basing a test entirely on orientation. Mr. Hugh Gordon, in a valuable series of experiments, has shown that the power to distinguish rapidly the right hand from the left, in any person and in any position,

is connected closely with general intelligence, and more closely still with mathematical ability. Hence the inclusion of dextrality questions among both the Chelsea Tests and the Picture Tests.

I have reason to think that this orientation test does in itself pick out children of extraordinary ability. One day when I happened to be applying the test in a boys' school the list of junior county scholarship winners arrived. The four boys who came out at the top in the orientation test were among the six who had gained scholarships.

In giving the Chelsea Tests the general procedure should be the same as that prescribed for the Columbian Tests. The children should first prepare an answer paper as represented in the Key.

The only test that requires special care on the part of the examiner is the cipher test, where his aim should be to ensure that the children know precisely what they are required to do without allowing enough practice for them to memorise the Key. Hence the preliminaries are fixed and standardised. Three minutes are allowed for the study of the instructions on the front page. Then the examiner explains by means of two examples what exactly is required. The point to be emphasised is that the pupil is not asked to translate the cipher (except of course in his own mind), but to write in ordinary script an answer to a question. The two blackboard examples I use are :

w-;c- ;s t-. f;rst d,y :f t-. w. .k ?
-:w m,ny l.tt.rs ,r. t-.r. ;n t-. w:rd c,t ?

The answers, I need scarcely say, are " Sunday " and " three."

After the expiration of the 10 minutes allowed for the cipher test the pupils are asked to fill up the blanks in that particular answer column with bold dashes to ensure that they are not filled in later. Then the pupils are told to work the rest of the tests entirely by themselves, keeping steadily on till they get to the end. No time limit is imposed.

The marking of the first three tests is a simple matter. One mark is allowed for each correct answer, and no fractional marks are given. In the fourth test the first three questions carry one mark each. For each of the other questions the score is:

(The number down right minus the number down wrong) multiplied by two.

Thus in Question 4—

if MVH is given, the score is	$(3 - 0)\,2 = 6$
if MV is given, the score is	$(2 - 0)\,2 = 4$
if MVHN is given, the score is	$(3 - 1)\,2 = 4$
if MHNS is given, the score is	$(2 - 2)\,2 = 0$
if MNS is given, the score is	$(1 - 2)\,2 = 0$

as no negative marks are recorded.

The possible marks therefore for this question are 6, 4, 2, and 0. The reason for this plan of marking is obvious. The "right minus wrong" principle counteracts the effect of guessing. The score is doubled because there are six judgments actually made, and the formula ensures that credit is given for each correct judgment.

Question 5 is marked on the same principle. The full score is 4, and the pupil may obtain 4 marks, 2 marks, or none.

Questions 6 and 7 are marked in the same way as Question 4.

The total marks for the third test are 25, as for each of the other tests, making an aggregate of 100.

I cannot with much confidence give age norms for the Chelsea Tests, for I have not yet applied them to a sufficient number of children. They are more difficult than the Columbian Tests, for while the average score in the Columbian Tests is roughly about five times the chronological age, the score in the Chelsea Tests is less than four times the chronological age. I submit the following as tentative norms:

Age	11	12	13	14	15	16
Norms	34	42	48	52	54	56

THE CHELSEA MENTAL TESTS

CIPHER TEST

(Three minutes are allowed for studying this page)

A printer once lost all his type for the five vowels and the letter "h," and he had to use punctuation marks for them as shown in the following key:

	a	e	i	o	u	h
KEY:	,	.	;	:	!	-

Instead of "a" he had to use a comma, instead of "e" a full stop, and so on. So, if he had to print the question, "How many eyes have you?" he would print it like this:

-:w m,ny .y.s -,v. y:! ?

There are 25 questions, printed in this funny way, on the other side of the paper, and you have to answer each of them by one word. If the above question were there, you would have to write down "two" on your answer paper.

Now try this one:

W-,t d: b,b;.s dr;nk ?

You will see that the answer is "milk."

Make no mark on the Question Paper.

Do not turn this paper over till you are told.

When you turn it over you will find the key on the top of the page.

You will be allowed only 10 minutes to answer the 25 questions, so you must work as fast as you can. Write quickly, but plainly, putting your answers in ordinary writing, not in cipher.

TEST I—CIPHER

Ten Minutes

Key: a e i o u h
, . ; : ! ·

Answer these questions :

1. C,n p;gs fly ?
2. W-,t c:l:!r ;s gr,ss ?
3. W-,t d: w. s.. w;t- ?
4. -:w m,ny l.gs -,s , d:g ?
5. ;n w-,t c:!ntry ;s P,r;s ?
6. W-,t s-,p. ;s , p.nny ?
7. W-,t ;s , s-;ll;ng m,d. :f ?
8. W-,t d: t-r.. ,nd tw: m,k. ?
9. D: d!cks w.,r w,;stc:,ts ?
10. W-,t ,r. b,by c,ts c,ll.d ?
11. N,m. t-. c,p;t,l :f .ngl,nd
12. ;s l.,d -.,vy :r l;g-t ?
13. N,m. t-. f;rst m.,l :f t-. d,y
14. W-,t s.,s:n c:m.s b.f:r. ,!t!mn ?
15. ;n w-,t c:!ntry d: t-. D!tc- l;v. ?
16. W-;c- ;s t-. sm,ll.r , c,m.l :r , :r:w ?
17. Wr;t. d:wn t-. t-;rd l.tt.r :f t-. ,lp-,b.t
18. W-,t ;s t-. n,m. :f :!r pr.s.nt k;ng ?
19. W-;c- ;s t-. l,rg.r , c:w :r ,n .l.p-,nt ?
20. -:w m,ny c;g,rs , d,y d:.s ,n .l.p-,nt sm:k ?
21. W-,t d,y :f t-. w..k c:m.s ,ft.r M:nd,y ?
22. W-.n , m,n -,s n: -,;r :n -;s -.,d w-,t d: w. s,y -. ;s ?
23. W-,t d:.s t-. m,n ;n t-. str..t p!t !p t: pr:t.ct -;m fr:m t-. r,;n ?
24. W-.n t-. -.,d :f t-. w.,t-.rc:ck p:;nts N:rt- ;n w-,t d;r.ct;:n d:.s ;ts t,;l p:;nt ?
25. ;f , m,n st,rts d;gg;ng ,t n::n w-,t t;m. w;ll ;t b. w-.n -. -,s w:rk.d f:r f;v. -:lrs ?

TEST II—MEANINGS OF WORDS

Look at this sentence: World, football, marble, melon are all (solid, eatable, round, small). There is one, and only one, of the 4 words in brackets that would make the sentence true. That word is "round." Now answer the first 8 questions by writing down for each one the word that makes the sentence true.

1. Honey, jam, saccharine, treacle are all (liquid, sweet, sticky, manufactured).
2. Diamond, platinum, radium, pearls are all (rare, jewels, transparent, round).
3. Greyhound, aeroplane, bullet, motor-car are all (metallic, self-acting, swift, munitions).
4. Grass, wood, apple, cactus are all (green, eatable, medicinal, vegetable).
5. Ant, grain, point, mite are all (hard, prickly, elongated, minute).
6. Gunpowder, knives, dynamite, bombs are all (chemicals, dangerous, explosive, cutlery).
7. Deserts, giraffes, rubber-soles, fishes are all (animals, thirsty, silent, spotted).
8. Birthdays, pain, error, death are all (disagreeable, inevitable, interchangeable, unconquerable).

Each of the following 6 questions is answered by one of the 6 phrases (A, B, C, D, E, F) given below. Answer each question by putting on your answer paper the letter that stands before the fitting phrase. (No words are required.)

9. Camel means ?
10. Punctuality means ?
11. Neck means ?
12. Patience means ?
13. Opportunity means ?
14. Crisis means ?

(A) Doing things at the exact time.
(B) A convenient time.
(C) An animal with a hump, used as a beast of burden.
(D) A time for deciding things, a decisive moment.
(E) The part of the body that joins the head to the trunk
(F) Waiting or suffering without grumbling.

Now do the next 7 questions in the same way. Be careful as some of the words have nearly the same meanings.

15. Authority means ?
16. Vanity means ?
17. Bravery means ?
18. Influence means ?
19. Pride means ?
20. Domination means ?
21. Fortitude means ?

a. A high opinion of one's own worth.
b. Rule that is exercised with force or with violence.
c. Power that rests on right.
d. Courage under suffering.
e. Self-esteem with a love of praise.
f. Force that comes with character or personality.
g. Courage under danger.

Now work the rest of the paper.

22. When the moon grows larger it waxes, when it gets smaller it ——. What is the missing word ?

23. When the tide comes in it flows, when it goes out it ——. What is the missing word ?

24. Which of these words means the largest number of things: houses, huts, towns, tents, dwellings, caves, cities ?

25. Which of these words means the largest number of things: cabs, barrows, trams, vehicles, carriages, omnibuses, carts, motor-cars ?

TEST III—ABSURDITIES

You will find below 25 statements each of which contains something silly; and after each statement there are four tries (A, B, C, D) at saying what is foolish in it. Read these carefully and find out which of the four is the best. If you think A is the best, write A on your answer paper. If you think B is the best, write B, and so on.

Look at this:

A soldier writing home to his mother said: "I am writing this letter with a sword in one hand and a pistol in the other." Foolish because—

A. The pistol might go off.
B. He could not write with a sword
C. He could not write with both hands occupied.
D. Perhaps his mother could not read.

The best reason is the third; therefore C should be put on the answer paper.

Work the following in the same way.

1. A boy who wanted to go to a cinema, but had no money, thought it would be a good plan to walk in backwards, for the man at the door would think he was going out, and would not ask for his ticket. Foolish because—

A. You cannot go in without a ticket.
B. The boy was trying to cheat.
C. The door-keeper would see that he was moving inwards.
D. Cinemas are bad for the eyes.

2. A countryman came up to London during the war, and saw on a hoarding the words: "Eat less bread: do it now." He immediately went into a tea-shop to do it now. Foolish because—

E. It said: "Eat less bread," not "Eat less buns."
F. He could not eat less bread by going into a tea-shop than by not going in.
G. Nobody ate more war bread than he could help.
H. If you eat too little bread, you will become ill.

3. An old gentleman complained that he could no longer walk round the park as he used to: he could now only go half-way round and back again. Foolish because—

K. It would be better to walk into the country.
L. The distances are the same.
M. He was getting lazy.
N. All old people are infirm.

4. A hunter who had used up all his ammunition was chased by a bear. A bright idea struck him: he would climb a tree. When he got to the top he remembered that the bear could also climb a tree; but he got out of his difficulty by pulling the tree up after him. Foolish because—

O. The tree could not be pulled up by the man if he himself was on it.
P. He was not strong enough.
Q. It was not a bright idea to climb a tree.
R. The bear would be too quick for him.

5. A forgetful man once tied a knot in his handkerchief to remind him of something; but he suddenly remembered that the last time he did this, he could not recall what he had to remember. So this time he tied two knots to make sure. The first knot was to remind him that he had to remember something, the second to remind him of what that something was. Foolish because—

S. It would spoil his handkerchief.
T. He should train his memory.
U. Two knots could not make him remember more than one.
V. He should have written it down.

6. Light comes from the sun; feathers are light; therefore feathers come from the sun. Foolish because—

W. The sun would burn the feathers.
X. The word "light" is used in two different senses
Y. Light comes from other things besides the sun.
Z. Feathers come from birds.

7. A sailor who was put to haul in a rope from the sea found it so long that he gave it up in disgust, saying that somebody must have cut the end off. Foolish because—

A. He should have had more patience.
B. All ropes have ends.
C. If the end were cut off, there would still be an end.
D. Ropes are used to keep the sails tight.

8. This is a sad and bitter world: we never strew flowers on a man's grave until he is dead. Foolish because—

E. We do not put a man in a grave before he is dead.
F. Only things we taste are bitter.
G. Dead men do not need flowers.
H. It is not the world that is sad, but the people in it.

9. The moon is more useful than the sun, for it gives us light in the night when we really need it, while the sun gives us light in the day when we don't need it. Foolish because—

K. When there is a moon the night is not dark.
L. The moon is not so bright as the sun.
M. On some nights there is no moon at all.
N. It is the sun that makes the day.

10. A householder saw an advertisement: "Buy one of Simpkins's stoves, and save half your coal." He bought two, to save all of it. Foolish because—

O. He should have bought more than two.
P. Two stoves would burn more than one.
Q. A stove wastes coal.
R. Simpkins was trying to swindle the public.

11. Some years ago it was proposed to shift Sunday to the middle of the week so as to divide the week into two parts. Foolish because—

S. There would still be seven days from Sunday to Sunday.
T. We cannot shift Sunday.
U. It is wicked to put Sunday where Wednesday is.
V Sunday must come after Saturday.

12. You are thin, and I am thin; but he is thinner than both of us put together. Foolish because—

W. We cannot make a thin man thinner.
X. It is better to be thin than stout.
Y. One man cannot be thinner than two men.
Z. If two men could be put together to make one man, that man would be fatter, not thinner.

13. Watching pigs wallowing in the mire, he muttered, "No wonder they are called pigs." Foolish because—

A. Pigs do not always wallow in the mire.
B. To call anybody a pig is not polite.
C. If a pig wallows at all, it must wallow in mire.
D. Pigs are not so named because they are dirty.

14. John James, who had married his widow's sister, used to say that if a man had a bad sister it was his misfortune, but if he had a bad wife it was his fault. Foolish because—

E. It is not lawful to marry one's widow's sister.
F. A good man may have a bad wife.
G. A woman does not become a widow until her husband is dead.
H. A bad wife is worse than a bad sister.

15. "I don't like onions," said the boy, "and I'm glad I don't, for if I did I should be eating them all day, and I hate the beastly things." Foolish because—

K. He cannot both like them and hate them at the same time.
L. No boy likes onions.
M. If he ate onions all day he would be ill.
N. Onions are not beastly.

16. An Irishman was charged at the police-court with having stolen a pig. Five witnesses swore that they had seen him do it; but the prisoner said that he could prove his innocence by bringing fifty witnesses who would swear that they had not seen him steal the pig. Foolish because—

O. A man would not steal a pig if people were watching him.
P. The people who did not see him were not there, so what they said did not matter.
Q. Witnesses do not always speak the truth.
R. Five witnesses were quite enough.

17. The three men laughed; then stopped suddenly as the eyes of each met those of the others across the table. Foolish because—

S. The three men might have been sitting on the same side of the table.
T. Looking would not make them stop laughing.
U. One pair of eyes cannot meet two other pairs at the same time.
V. Eyes do not meet; they look.

18. I am not conceited, for I don't think I am half as clever as I really am. Foolish because—

W. He is not so clever as he thinks he is.
X. He says he thinks he is clever and not clever at the same time.
Y. He can be clever without being conceited.
Z. A man should not brag about his cleverness.

19. A man who bought a dog that had been advertised complained to the seller that the dog's legs were too short. The seller replied, "They are long enough to reach the ground, aren't they? What more do you want?" Foolish because—

A. Some kinds of dogs always have short legs.
B. A dog's legs may reach the ground, and yet be both ugly and useless.
C. We cannot make a dog's legs longer by stretching them.
D. Short legs are often as pretty as long ones.

20. One should never use long words when short ones will do just as well, for when the words are long the meaning is obscured by the unusualness of the terminology, and intelligibility is consequently endangered. Foolish because—

E. Long words sound better than short ones.
F. The man who said this did not follow his own advice.
G. Short words are easier to spell.
H. It does not matter whether words are long or short.

21. "I have always noticed," remarked the old man, "that if I

do not die in March I am all right for the rest of the year." Foolish because—

K. March is not so cold as January.
L. Nobody knows when he is going to die.
M. He may die any time.
N. He might have said the same thing of any other month of the year.

22. A showman advertised for a giant and a dwarf. A man of ordinary height presented himself and offered to fill both parts. He claimed to be the smallest giant in the world, and the biggest dwarf. Foolish because—

O. He wanted to cheat the showman.
P. No man can be two things at once.
Q. A giant must be larger and a dwarf smaller than an ordinary man.
R. The showman advertised for two men, not one.

23. Every rule, even this one, has an exception. Foolish because—

S. This is not a rule.
T. Some rules have no exception.
U. It contradicts itself.
V. We cannot say which is the exception.

24. While standing near a clock-tower just before the clock struck twelve two boys tried to find out which of them could hold his breath the longer. Neither of them won, for one was able to hold his breath from the first stroke of the clock to the sixth, and the other from the sixth to the twelfth. Foolish because—

W. The boys could not hold their breath so long.
X. It is unhealthy to stop breathing.
Y. They ought to have started together.
Z. The second boy won.

25. The horse obeys his master because his eyes magnify, so that his master seems to the horse to be much larger than the horse himself. Foolish because—

A. A man's eyes magnify too.
B. We cannot tell whether a horse's eyes magnify.
C. The horse obeys his master because he respects him.
D. The horse's eyes would magnify everything in the same proportion.

TEST IV—ORIENTATION

Keep your paper straight in front of you.

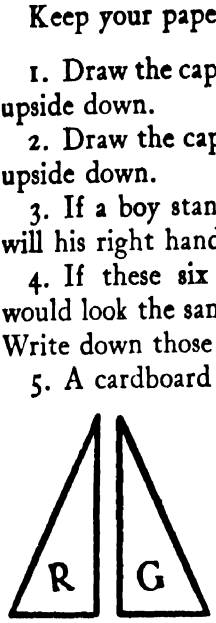
1. Draw the capital letter **F** as it would appear if it were printed upside down.

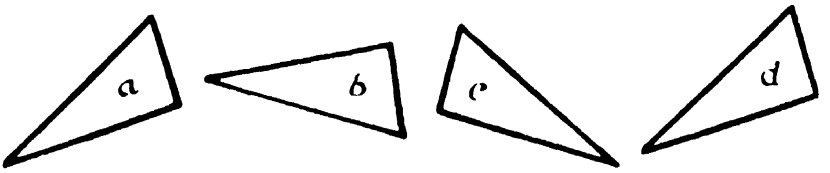
2. Draw the capital letter **N** as it would appear if it were printed upside down.

3. If a boy stands on his head with his face to the south, where will his right hand point?

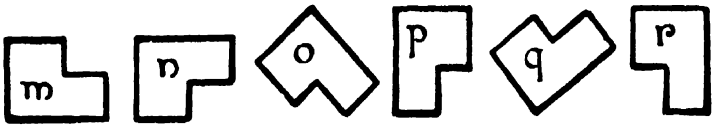
4. If these six letters were seen reflected in a mirror, some would look the same as they do here and some would look different. Write down those that would look the same: **SPMVNH.**

5. A cardboard triangle is painted red on one side and green on the other. The drawing marked **R** shows what it looks like when it lies with the red side up, and the drawing marked **G** shows what it looks like when it is turned over so that the green side is up. This cardboard was placed on this paper in four positions named *a*, *b*, *c*, and *d*, sometimes with the red side up and sometimes with the green side up. In which positions was it when the red was up?

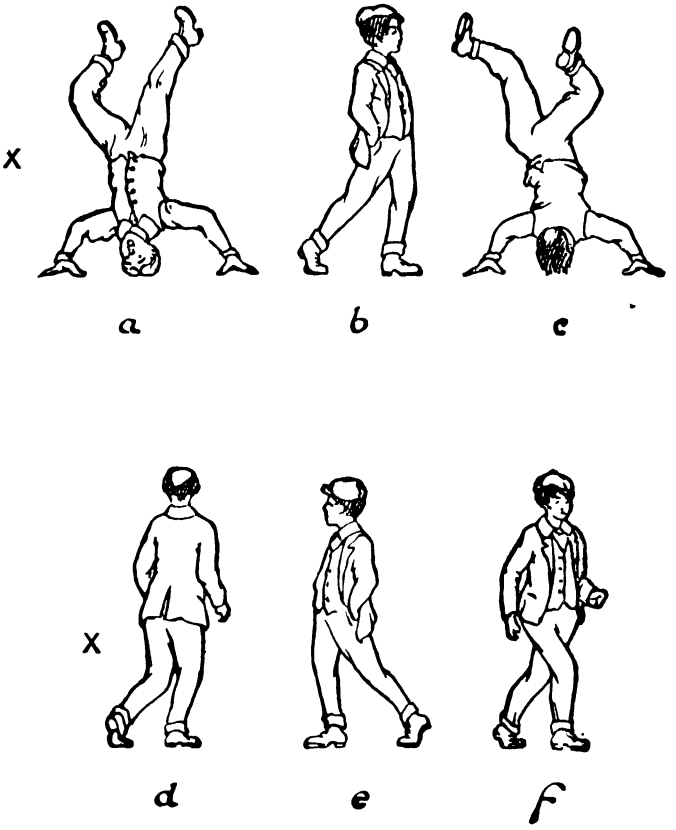




6. This question is like the last one, except that the cardboard is of a different shape and I give you a drawing of the green side only. The red side is underneath. In which of the positions *m*, *n*, *o*, *p*, *q*, and *r* was the red side up?



7. Name by letters all the people you see drawn here whose left leg is nearer the beginning of the line (nearer the **X**) than the right leg.



THE CHELSEA MENTAL TESTS

KEY

No	Test i.	Test ii.	Test iii.	Test iv.
1	No	Sweet	C	**d**
2	Green	Rare	F	**N**
3	Eyes	Swift	L	East
4	Four	Vegetable	O	**M V H** (6 marks)
5	France	Minute	U	*b*, *c* (4 marks)
6	Round	Dangerous	X	*m*, *p*, *q* (6 marks)
7	Silver	Silent	C	*a*, *d*, *f* (6 marks)
8	Five	Inevitable	E	
9	No	C	N	
10	Kittens	A	P	
11	London	E	S	
12	Heavy	F	Z	
13	Breakfast	B	D	
14	Summer	D	G	
15	Holland	*c*	K	
16	Crow	*e*	P	
17	C	*g*	U	
18	George	*f*	X	
19	Elephant	*a*	B	
20	None	*b*	F	
21	Tuesday	*d*	N	
22	Bald	wanes	Q	
23	Umbrella	ebbs	U	
24	South	Dwellings	Z	
25	Five o'clock	Vehicles	D	

CHAPTER XV

THE CRICHTON TEST

For the more intelligent adolescents and adults I have devised a test which may be used to supplement the Chelsea Tests. It serves to differentiate still further the higher grades of intelligence. I have called it the Crichton Test, as the child who could answer all the questions would bid fair to rival the Admirable Crichton. The type of test was suggested by Roback's Mentality Tests for Superior Adults, which end up with a cryptogram of nine words. As the number of symbols used is only eleven, and five of them are already translated, it will be seen to be easier than the two cryptograms included in the present series of tests. The data I have supplied for their solution consist of certain principles rather than partial translation.

No preliminary instructions are necessary. The papers are distributed, and the candidates told to work them as though they were ordinary examination papers. No mark is to be made on the question paper, but the candidate is allowed—nay, recommended—to use the back of his answer paper for his attempts at deciphering the two cryptograms.

The passage in No. 6 is from Marcus Aurelius, in No. 7 from Kenneth Graham's *Pagan Papers*, in No. 8 from Emerson, in No. 9 from Oliver Wendell

Holmes, in No. 10 from Emerson, in No. 11 from Hilaire Belloc's *Caliban's Guide to Letters*, and No. 12 from Boswell's *Johnson.*

The first twenty-three questions carry 1 mark each, the twenty-fourth carries 2, the twenty-fifth 3, the twenty-sixth 3, the twenty-seventh 8 (one for each word), and the twenty-ninth 11 (one for each word). The total is 50.

I have not yet standardized the results of the Crichton Test, but some notion of its difficulty may be gained from the fact that the average score for the top class of an elementary school (Standard VII) is about 7, and there are never more than one or two who are able to solve the last problem. At a large Central School for boys, the 42 brightest pupils (about one-ninth of the school) were given the test. The average score was 23, and the highest 43. The second and harder cryptogram was deciphered by 17 boys.

THE CRICHTON TEST

In each of the first 5 questions you have to find a fourth word that goes with the third as the second goes with the first.

Sample: Table is to *wood* as *window* is to —— ? Ans.: glass.

1. *Music* is to *opera* as *poetry* is to —— ?
2. *Word* is to *thing* as *money* is to —— ?
3. *Tub* is to *but* as *pot* is to —— ?
4. *House* is to *furniture* as *mind* is to —— ?
5. *Aspect* is to *space* as *phase* is to —— ?

In each of the next 4 passages one word has been altered so as to make the meaning nonsensical. In each passage pick out the misfitting word.

6. The mind converts and changes every hindrance into help. And thus it is probable I may gain by opposition and let the obstacle hinder me on the road.

7. In book-buying you not infrequently condone an extravagance by the reflection that this particular purchase will be a bad investment: that you are not squandering income but sinking capital.

8. The exclusive in fashionable life does not see that he excludes himself from enjoyment in the attempt to appropriate it. The exclusionist in religion does not see that he shuts the door of heaven on himself, in striving to shut in others.

9. The heart is a great reservoir, which distributes food, drink, air, and heat from every part of the system, in exchange for its waste material. It knocks at the gate of every organ seventy or eighty times a minute, calling upon it to receive its supplies and unload its refuse.

Among the following 8 sentences some are witty (call them W); others are wise as well and express profound truths (call them T); others are commonplace (call them C); and others are nonsensical, as they either say something silly or say the same thing twice (call them N). Mark each on your answer paper as W, T, C, or N.

10. If you put a chain around the neck of a slave the other end fastens itself around your own.

11. Show me the man hour by hour in his own home, from the rising of the sun to its going down, and I will tell you what manner of man he is.

12. No man will be a sailor who has contrivance enough to get himself into a jail; for, being in a ship is being in a jail with the chance of being drowned.

13. A gentleman is a man who is never rude unintentionally.

14. Our virtues are like trees; they extend further and further as they grow larger.

15. One man with truth on his side constitutes a majority.

16. The man who raises his hand against a woman except in the way of kindness is a contemptible scoundrel.

17. If a man begins to think he is as wise as he is good, he will end by thinking that he is as good as he is wise.

18. If I toss up two pennies together several times they may come down in 3 different ways, *i.e.* as H H, H T, or T T (where H means head and T tail). In how many different ways may four pennies come down?

19. In a certain family there are four boys who are allowed to go to the swimming baths two at a time. In how many different pairs could they go?

Several little wooden cubes each with an edge of one inch are put together to form a solid cube with an edge of three inches. This big cube is then painted red all over the outside. When the paint is dry the big cube is broken up into the original little cubes. Now answer these four questions about the little cubes:

20. How many cubes will have paint on four sides?
21. „ „ „ „ „ „ „ three „ ?
22. „ „ „ „ „ „ „ two „ ?
23. „ „ „ „ „ no paint on ?

24. Read question 18 again and then say in how many different ways 100 pennies may come down if tossed into the air together? (2 marks.)

25. The first odd number is 1; the second odd number is 3; the third odd number is 5, and so on. What is the 200th odd number? (3 marks.)

26. If 16 oranges are arranged on the ground in a compact square with four forming each side of the square, and then more oranges are placed on top to form a pyramid, that is a heap getting regularly less as it rises, until it ends with only one on top. How many oranges are there in this pyramid? (3 marks.)

In dealing with the next two questions you may scribble on the back of your answer paper.

27. The sentence given below is printed in cipher: figures are put to represent letters. Translate it. The following facts will help you:

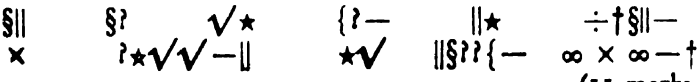
(*a*) There are in the English language very few words of one letter.

(*b*) The letter most frequently occurring at the end of words is "e."

(*c*) 3 stands for "l" and 4 for "t."

2 870 4574 2 078 7 324431 3741 (8 marks.)

28. The following symbols make up a sentence. Translate it bearing in mind what was said under (*a*) and (*b*) in the previous question, and noting that there are in English only a few common words of two letters, and that there is a certain pair of two-lettered words one of which is the same as the other spelled backwards



(11 marks.)

THE CRICHTON TEST

KEY

(1) Drama (plays). (2) wealth (goods, commodities). (3) top. (4) knowledge (ideas, thoughts). (5) time. (6) hinder. (7) bad. (8) in. (9) from. (10) T. (11) C. (12) W. (13) W. (14) N. (15) T. (16) C. (17) N. (18) 5. (19) 6. (20) 0. (21) 8. (22) 12. (23) 1. (24) 101 (2 marks). (25) 399 (3 marks). (26) 30 (3 marks). (27) I saw that I was a little late (8 marks). (28) It is no use to write a sonnet on tissue paper (11 marks).

CHAPTER XVI

WHAT IS INTELLIGENCE?

THERE are many ways of approaching this question. One way is to examine the concept as it exists in the minds of the majority of people who use the term, or to examine the sense in which the term is used by modern psychologists. A quest of this kind is not without profit: it at least serves to clarify the mind of the searcher.

The Editor of *The Journal of Educational Psychology* recently invited a number of leading investigators in America to state what they meant by "Intelligence," and the replies sent in by thirteen of them appear in the issues of the *Journal* for March and April 1921. These replies reveal a surprising variety of opinions: no common nucleus of meaning is discoverable. Nor do the views of European psychologists show much more unanimity. It is possible, however, to group the various opinions under three heads: those that regard intelligence as a single ability common to all intellectual processes; those that regard it as a group of two or three abilities of varying degrees of generality; and those that regard it as representing no real entity but as merely a convenient term for the average of all specific abilities.

Of those writers who hold the first theory many

accept Stern's[1] definition: "*General adaptability* to the new problems and conditions of life." Burt[2] defines intelligence as "inborn, all-round mental efficiency." Woodrow[3] calls it "a capacity to acquire capacity"—a phrase that suggests an endless regression. Spearman,[4] whose name is most closely connected with this theory, prefers calling the central factor "general ability" (symbolised by *g*) without committing himself to the view that *g* is identical with intelligence as commonly understood.

The second theory, the group theory, is held by Binet,[5] who regards intelligence as involving at least three factors, which are thus summarised by Terman[6]: "(1) Its tendency to take and maintain a definite direction; (2) the capacity to make adaptations for the purpose of attaining a desired end; and (3) the power of auto-criticism." Another interesting view of the composite nature of intelligence is suggested by Maxwell Garnett,[7] who says: "It is probable that genius, as the word is commonly understood, is more directly measured by $\epsilon = \sqrt{g^2 + c^2}$ than by *g* (measuring General Ability or capacity to concentrate attention) alone, or even by *c* (measuring Cleverness, as we have defined it, or tendency to associate by similarity) alone." In other words, the general factor *g* is

[1] *J. of Educ. Psychol.*, xii. 127.
[2] *J. of Exper. Ped.*, i. 95.
[3] *J. of Educ. Psychol.*, xii. 208.
[4] *British Journal of Psychol.*, v. 51–84.
[5] "L'intelligence des imbéciles," in *L'Année Psychologique* (1909), 1–147.
[6] *The Measurement of Intelligence*, p. 45.
[7] *Education and World Citizenship*, 125.

compounded with one, or perhaps more than one, group factor to constitute Intelligence.

The third theory, the theory of the independence of the several traits that go to make up the mind, has been advocated with wavering confidence by Thorndike.[1] Though, in his latest pronouncement, he admits a certain possibility of prophesying from a test score "how well a person will do in other intellectual tasks," he warns us that "to assume that we have measured some general power which resides in him and determines his ability in every variety of intellectual task in its entirety is to fly directly in the face of all that is known about the organisation of intellect." Under the same head comes G. H. Thomson's[2] Sampling Theory of Ability, which assumes that each man's mind is a sample group of qualities, resembling Mendelian units, which come from the hereditary stock. "If there be a General Factor at all, it might be the power to shake down rapidly into good team work."

In spite of this divergence and apparent inconsistency of opinion, many of the individual views seem merely to accentuate one or more aspects of a complex whole. At any rate the majority of the theorists would probably subscribe to the following tenets: Intelligence is general mental ability which operates in many different ways; it is more fully manifested in the higher mental processes than in the lower; it is specially active in dealing with a situation which presents points of novelty (in other words with the solution of problems); it is more

[1] *J. of Educ. Psychol.*, xii. 126.

[2] *Essentials of Mental Measurement*, by Brown and Thomson, 1921, 188–92

concerned with the dissecting, planning, and rearranging of the data of experience than with the mere reception of impressions.

It does not seem to me to be possible to maintain a clear-cut distinction between powers that are innate and powers that are acquired, and to say that intelligence is concerned with the former and not with the latter. For the truth is that every type of mental ability, whatever its degree of generality or particularity, is both innate and acquired: it is born with a man as a potentiality, it is acquired as an actuality. The most instinctive of man's tendencies requires some degree of stimulus from nature's schoolmaster—the physical and social environment—before it can operate at all; and some amount of exercise and schooling before it can operate efficiently. In other words, it needs training. Short and rapid and unconscious though the process may be, it is a training process just the same. The innate predisposition has become an acquired ability. On the other hand, every acquired ability was at one time an innate predisposition—a latent or potential ability. Something was there from the first, whether that something was merely neural or psychical or both

The ability to utter articulate speech is generally recognised as an acquired ability: without the schooling of special circumstance it cannot exist. The man reared among the beasts of the forest (like that ineffectual Tarzan, the wild boy of Aveyron) would be as inarticulate as the beasts themselves; but in their relationship towards language there would be a world of difference between the mind of the man and the mind of the brute. There would

be a potentiality in the one which would be absent in the other. Neither the man nor the brute would possess *de facto* the power of speech; but one would have the capacity to acquire this power, while the other would not. All that would be lacking to make the difference actual and tangible would be the stimulus of a human environment. It would sting into action the slumbering faculty of the man: it would beat in vain on the mind of the brute.

The world of wild nature which we share with beast and bird is in itself insufficient to awaken the faculty of speech: the narrower and richer world of civilised life which we share with our human brothers, while sufficient to ensure articulate speech, is insufficient to bring into being many other latent human powers, such as the power to play the violin or the power to solve quadratic equations. Here the milieu has to be narrowed and specialised still further. It has to become a stage-managed environment. Then only will the capacity to do these special things—using the term "capacity" in the sense of latent or potential ability—be turned into the ability to do these special things. But the capacity is inborn; and by a slight extension of meaning we may also say that the ability is inborn.

Thus the distinction between innate and acquired ability is relative, not absolute; and it is a distinction which rests on two factors: first the facility with which the appropriate environment can rouse that ability to action (in other words the ease with which it can be trained), and secondly the degree of speciality in the environment itself. The shorter the period of training and the more general and casual the environment, the more inclined are we to

call the ability innate; the longer the period of training and the more special and sophisticated the environment, the more disposed are we to call the ability acquired. Relatively to one another the ability to see is innate and the ability to speak acquired; relatively to one another the ability to speak is innate and the ability to read acquired.

Nor can we differentiate the two types by saying that inborn abilities have a limit of trainability while acquired abilities have not. Both have their limits. Both are trained along the same lines and pass through the same stages of progress. Practice produces rapid improvement at first, then the rate of improvement gradually slows down, and finally stops altogether. All that further practice can do is to prevent deterioration. The curve of development, which begins by rising steeply, gradually bends towards the horizontal, and ultimately ceases to rise at all. The force expended in training is subject to what economists call "the law of diminishing returns." The process of training a mental power seems to me to resemble in some respects (the simile must not be pressed too far) the drawing out of a spiral spring fixed at one end. The spiral yields easily at first, but the straighter and longer it gets the greater the force required to stretch it, until at last when it gets quite straight it can be stretched no further. But force is still needed to keep it straight. Education needs a special and organised attempt to pull out the spirals. When they seem to come out of their own accord (they do not really) the corresponding abilities are said to be innate.

So far we have viewed the problem of intelligence

from the side of the preconceived concept. Let us now approach it from the other side. Let us examine the tests themselves and see what light they throw upon the problem. The tests purport to measure some mental factor or function—whether simple or composite does not matter—and it is the general opinion that, roughly speaking, they succeed. They really do measure something. What is the nature of that something? There are some at least of its qualities which are clearly defined. It grows with a child's growth up to a certain age and then appears to stop; it seems to bear, up to about sixteen, a constant ratio to chronological age; it is independent of special schooling or special training; and it is closely related to success in school and success in life. And if the apostles of the *a priori* protest and say that that is not what we mean by intelligence, we can only reply that it is what the mental tester means by intelligence, and it is what the mental tester means by intelligence that we are trying to expound. It may well be that he had no right to use the term in this sense; but he borrowed it for lack of a better, and in the very attempt to measure it his notion of its nature underwent a change. So it came to pass that what the modern psychologist means by intelligence is very different from what the arm-chair philosopher means by intelligence. Let us then examine this new concept a little more closely; and having criticised the tests in the light of the old concept, let us criticise the old concept in the light of the tests.

There is nothing new in the idea of measuring a thing first and finding out what it is afterwards. We do not yet know what electricity is, but we are

so sure that it can be measured that we pay our electric-light bill without a protest.

The merest glance at the armoury of mental tests will suffice to show that they do not eliminate knowledge. There is not a single test in Binet's scale, or in any other scale, which does not assume some amount of knowledge on the part of the subject. Indeed what the tests test is often nothing but knowledge. When the Binet tester says to a child of nine, "Name the months of the year"; or to a child of six, "Show me your right hand"; he is seeking evidence of knowledge as a personal possession. When the American soldier was asked to state whether denim was a dance, a food, a fabric, or a drink, he was asked to exhibit some of the knowledge he had amassed. It is clear that intelligence cannot be tested *in vacuo*: it can only be tested in relation to knowledge. We must find out how it deals with knowledge—how it acquires it, at what rate it acquires it, and what it can do with it after acquiring it. Given two minds similarly circumstanced (a difficult if not impossible condition to secure), the mind that absorbs knowledge the more rapidly is the more intelligent of the two—other things being equal. Among these other things the most important is the nature of the knowledge that is absorbed—whether it is the knowledge that is power or the knowledge that lies as a dead weight upon the mind.

If, then, both mental tests and examinations inquire into the knowledge possessed by the subject, what difference is there between them? There is this difference. The knowledge that the tester tests is knowledge that the subject of ordinary intelligence cannot possibly avoid. He does not seek it, it

seeks him; it thrusts itself upon him in the common intercourse of daily life. The examiner, on the other hand, tests knowledge that the candidate can all too easily avoid (and often does). If he wants it he must seek it in curious places—in books and schools and laboratories. One type of inquiry postulates a common background of experience in order to gauge ability; the other postulates nothing: it merely estimates attainments.

Although the mere possession of knowledge is presumptive evidence of intelligence, the power to use that knowledge is evidence of a much stronger kind. Indeed the power to apply old knowledge to new situations has often been put forward as the sole criterion of intelligence. The commonest definitions of the term contain as a nucleus the notion of adjustment, of adaptation, of conforming to new conditions, or, as some prefer to put it, of solving problems. This is intelligence in the narrower sense, intelligence in its higher manifestations—at least if we give the usual meaning to the term problem. And in this sense of the term it is far too narrow to cover the whole field of mental testing. If intelligence means the ability to solve problems, then the American group tests are not tests of intelligence. Although therefore in both the mind of the plain man and the mind of the psychologist the concept of intelligence is undoubtedly biased towards what Dr. Maxwell Garnett calls "cleverness," cleverness is not regarded by the mental tester as an essential part of the thing he is trying to measure.

It is commonly agreed that the facilitation that comes from usage and habit is excluded from what we

mean by intelligence. The man who can add up a column of figures twice as fast as another is not *twice* as intelligent—not necessarily more intelligent at all. He has probably had more practice, that is all. The *computation has become more automatic*; and automatism is the very antithesis of intelligence. Yet from the very fact that all the American group tests are time tests we see that it is just as impossible to eliminate habituation from mental testing as it is to eliminate knowledge.

The upshot of our discussion is that if we generally accept modern mental tests, not of course as perfect instruments of inquiry, but as instruments of inquiry that are right in kind and reasonably workable, then we must broaden our concept of intelligence, and we must face the fact that we cannot in our actual testing eliminate knowledge, interest, and habituation: the most we can do is, in a very rough and ready way, to standardise their influence. We must broaden our concept of intelligence; for it is difficult to find any form of mental activity which intelligence tests do not gauge, any corner of the mind which they do not probe. Nothing short of the mind's general efficiency in dealing with experience will meet the case. And here we see why it is that English psychologists have as a rule avoided the word "intelligence" and have used instead the broader term "general ability"; and here too, if we adopt that wide interpretation of "inborn" which we have given above, and which Mr. Burt as a matter of fact accepts, we see the fitness of Mr. Burt's definition of intelligence as "inborn, all-round mental efficiency." Professor Adams too has, in his happy way, expressed the same view in calling

intelligence "applied thought"—thought at work in the world of experience.

Venturing still farther along the thorny path of definition, I may say that intelligence is the relative general efficiency of minds measured under similar conditions of knowledge, interest, and habituation.

CHAPTER XVII

THE LIMITS OF INTELLIGENCE

If we accept the findings of the last chapter we are brought face to face with two startling characteristics. The first is that intelligence is unaffected by schooling, and the second that it ceases to grow beyond the middle of the teens. Not only are the tests themselves so framed as to exclude book-learning, but the training a child receives at school gives him no advantage over the illiterate in coping with these tests. This is capable of proof. Many a child has been tested, and his intelligence quotient evaluated, before he has entered school. That quotient (the ratio his mental age bears to his real age) has remained virtually unchanged during the whole of his school career.

These broad statements—made broad for the sake of clearness—stand in need of some slight qualifications. For they are really conclusions towards which the evidence is gradually forcing us rather than truths capable of strict mathematical proof. It is not quite true that the tests as they stand at present entirely ignore book-learning; but rather that that is the ideal towards which they are tending. It is not quite true that the intelligence quotient always remains constant, but it is so often found to be nearly constant that the deviations from the rule,

when not due to imperfections of measurement, are probably due to accidents of health, nutrition, or fatigue. It is perhaps possible that the development of intelligence is hastened by education; but if so the acceleration is so slight as to be negligible. It is only then in a broad and general sense that it can be maintained that the mental factor that is being tested is independent of schooling and unsusceptible of training.

From the fact that intelligence cannot be trained we must beware of drawing false conclusions: we must particularly beware of inferring that intelligence does not concern the teacher. The truth is that it vitally concerns him. The intelligence of his pupils is his most valuable asset; nothing else in the classroom, neither its position nor its equipment, nor the age of his pupils nor their social status, nor their previous training, nor anything else, is so essential to success in the exercise of his craft. There is all the difference in the world between intelligent teaching and mechanical teaching, as there is all the difference in the world between intelligent learning and learning by rote, by rule of thumb, by imitation, by trial and error, or by any other method that shirks the need for thinking. Though the teacher cannot train intelligence he can use it. Though it is not the material *on* which he works, it is the material *with* which he works. We have Thring's authority for saying that the boy who does not use the sense he has got is a fool. It is, then, the teacher's business to see that his boys do not make fools of themselves.

When you tell the plain man that intelligence is untrainable he finds no difficulty in accepting the

doctrine. He is quite willing to believe that he owes his intelligence to himself, or his ancestors, and not to his teachers. But when you tell him that he is no more intelligent now than he was at 16 he will think you are talking paradox, which is one degree worse than talking nonsense. And when you assure him that you are in earnest he becomes angry and says he was a young ignoramus at sixteen—which is no doubt true, but beside the point. We must admit that not only is the notion galling to our pride, but it sometimes seems to us that all that we know that is worth knowing, all our real culture and erudition, have been acquired since the age of 16; and in the teeth of this strong conviction it seems preposterous to assert that our intelligence has not only failed to keep pace with our culture, but has remained exactly where it was. And it is on our intelligence that we really pride ourselves: we feel more sure of it than of our erudition. This feeling is rooted deep not only in the public mind, but in the professional mind. Let us then look calmly at the evidence. We first note that Binet's scale of mental ages did not extend beyond 15. It is true that he added a group of five tests for adults, but these tests have since been found to be too difficult for the average adult. They fail to fit in with the rest of Binet's scale, which is strictly based on age-performance experimentally determined. Instead of indicating a norm they fix an arbitrary standard. As, however, Binet was primarily concerned with the detection of feeble-minded children it was but natural that he should have standardised the lower part of his scale more carefully than the upper part.

In making the Stanford Revision of the Binet

Scale Terman[1] was influenced by no such restriction of purpose, and yet he could carry his age scale no further than 14. To represent the higher ages he has two sets of tests, one for the "average adult" and the other for the "superior adult." In calculating the intelligence quotient he assumes the chronological age of all persons over 16 to be 16; which is equivalent to assuming that after 16 years of age there is no improvement in native intelligence.

Burt does not carry his scale of reasoning tests beyond 14.

These facts suggest either an early cessation of development or a difficulty in devising tests for mental ages beyond 15. My own belief inclined to the latter alternative; I was reluctant to accept the former without some sort of verification I therefore set about putting the matter to the test of experiment in the hope of finding the psychologists wrong and of showing that there was an improvement either in quantity or kind which they had failed to detect. And when I asked myself what kind of improvement one would naturally expect the answer seemed to be an improvement in reasoning and common sense; and as a test of reasoning and common sense I used the absurdities described in Chapter VII.

The main obstacle in the way of research is the difficulty of securing unselected subjects over 14 years of age. Up to 14 years of age there is but little difficulty; for all children between 5 and 14 are compelled by law to attend school, where they can be found and tested. But after 14 in almost every type of school the duller pupils begin to leave; and

[1] *The Measurement of Intelligence*, 56–61.

as the ages increase the age-groups get increasingly smaller, and the quality of intellect increasingly higher. Such improvement therefore as one would find in the intelligence of secondary school children as they proceed up the school is not necessarily due to the growth of intelligence with age: it may be due to the gradual elimination of the less intelligent.

It is almost impossible to feel certain that one has secured a good representative sample of any age-group beyond that which falls between 13 and 14; and the further the group is removed from this the more difficult becomes the task.

I applied the 34 absurdity tests to about 2,000 pupils in a variety of schools—wherever in fact I could find a reasonable number of pupils ranging between the ages of 11 and 18—and I found that the adults at a woman's training college did no better than the girls of 16 at secondary schools, and the adults at an evening school did much worse. In some of the schools tested there was no evidence of improvement after 12 years of age. Where, as at some secondary schools, there was a full range of ages from 11 to 18, little improvement was shown after 15 and none at all after 16. Thus I was forced to conclude that the psychologists were probably right after all.

I could wish that the evidence I secured were stronger; that the subjects tested had been more homogeneous, and that the tests had been more numerous and more varied. It must be admitted too that the validity of my results rests on the assumption that the successive age-groups tested were on the average of the same intelligence quotient—that the fifteen-year-old subjects, say, would be in

a year's time exactly as intelligent as the sixteen-year-old subjects were at the time of testing. Only thus could we regard the curve determined by the aggregate results as indicating the normal growth of intelligence.

If these results stood by themselves, they would

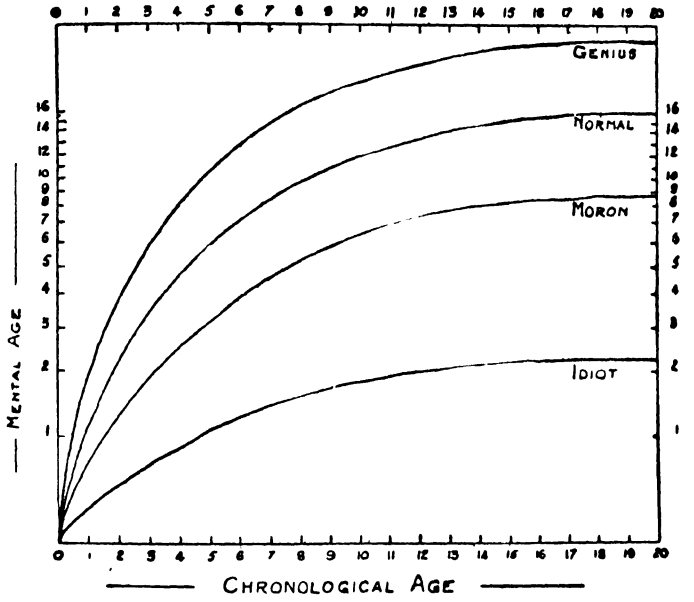


FIG. 18.—DR. WOODROW'S DIAGRAM.

have to be received with extreme caution. As, however, they confirm the findings of other psychologists, we need have less hesitation in accepting them. Such evidence as we possess concerning the growth of intelligence forces us to believe that Dr. Woodrow's diagram (Fig. 18) gives a fairly accurate representation of the essential facts. We may state

generally that the growth of intelligence, which is rapid for the first few years, gradually slows down as the child gets older, that after 12 years of age the slowing-down is very marked, and that after 16 farther growth is inappreciable. It is this retardation of growth that accounts for the fact that Binet failed to interpolate in his scale a set of tests either between the ages of 10 and 12 or between the ages of 12 and 15; omitting the doubtful tests for adults he could only mark out two steps after the age of 10.

It follows that a year of mental age is not a fixed unit; it gradually diminishes towards the higher end of the scale.

If after 16 years of age the curve of growth is virtually horizontal, as it seems to be, we are forced to conclude that 16 is not the average age at which growth ceases, but the maximum age. For the general curve is the resultant of a large number of individual curves; and if the resultant curve is horizontal after 16, then either all the individual curves are horizontal after 16, or if some rise others must fall. In other words, we must choose between two alternative suppositions, one that nobody improves after 16, the other that some people in sound health begin at that age to suffer from senile decay. There is evidence for thinking that different races and different individuals vary in the age at which mental maturity is reached. S. D. Porteus [1] found that Australian aborigines complete their growth in intelligence several years earlier than white children. The average mental age of recruits in the American army during the war is said to be somewhere be-

[1] "Mental Tests with Delinquents and Australian Aboriginal Children," *Psychol, Rev.*, xxiv. 32.

tween 13 and 14. It is highly probable, therefore, that the average age at which growth ceases is considerably below 16. The exact determination of this average age is of importance in estimating the intelligence quotient of adults; for if, following Terman, we use 16 as the denominator for all chronological ages over 16, we undoubtedly get a quotient which is too low.

There is obviously need for an investigation which aims at finding the separate curves of growth for a large number of individuals, an investigation that will be wide in its scope and will be carried over a long series of years. It is only by an extensive research of this kind that we can find an answer to the question: What connection is there between the quality of a person's mind and the age at which his intelligence reaches maturity? Are "late bloomers" better than "early bloomers"?

Of those who reject the doctrine of intelligence as here expounded none has more ably expressed the grounds of his rejection than Professor J. A. Green,[1] of Sheffield. His thesis is that intelligence cannot be separated from the universe within which it works; as a man's universe changes so does his intelligence change, when his universe expands his intelligence expands as well. Each conditions the other; and intelligence is revealed not by a sup-

[1] Since writing this chapter the sad news has reached me of Professor Green's death. My first impulse was to delete what I had here written in reply to his recent attacks on the current doctrine of intelligence. On second thoughts I decided to let it stand, feeling that he himself would not wish it otherwise. Although we took opposing sides in this controversy, he was personally my friend, and a friend for whose ability and whose services to our common cause I had abundant admiration.

posed activity *in abstracto*, but by its universe of operation. To test the intelligence of adults by absurdities is itself an absurdity.

To measure the intelligence of a man of 40 by the same tests and the same scale as are used in measuring the intelligence of a youth of 16 is to ignore an essential part of the problem—the nature and extent of the effective environment—and to place the man under a serious disadvantage.

That, I take it, is the gist of Professor Green's argument, an argument which gains plausibility from the fact that increased experience brings with it increased power—that knowledge may, as Herbart has abundantly shown, become itself a kind of ability; or, to use a Spencerian phrase, that fact may be changed into faculty. The argument gains plausibility too from the wider interests that come with advancing years. The world that interests us at 40 is much larger than the world that interests us at 16. And community of interests is often mistaken for keenness of intelligence. We are disposed to form a high estimate of the man who is interested in what we are interested in, who laughs at our jokes, who accepts without demur our opinions and our prejudices, the bee in whose bonnet buzzes in tune with the bee in our own. Then, again, there is the facility which custom and habit bring to all human activities. As we grow older we feel more at home in the world: we do things with less expenditure of attention and effort; we have formed thought-saving habits, and what thought we do bring to bear upon our business or our pleasure is used with greater economy and effectiveness.

All this is so obvious that it is mentioned only to

point out that it has nothing whatever to do with intelligence: it is altogether extraneous to the mind's inherent and original power of applying thought to the world of experience. A man with a lever can exert more force than a man who uses brute muscle, and a man with a long lever can exert more force than a man with a short lever. But if we wish to compare the men on the basis of their muscular strength we must take away their levers; or if this is impossible, we must give them all levers of the same size. This is precisely what we try to do when we test intelligence. We cannot take away all the instruments which the mind has forged for itself in its commerce with the world, but we can see to it that the instruments used are of approximately the same power; we can deprive the older subjects of the advantage gained by more extensive knowledge and more facile skill. We can make them all "start at scratch." There is no injustice in this kind of handicap: it is the scientific means of eliminating the influence of factors which form no part of the essence of the problem.

Another argument that has been used is that intelligence is an "abstraction." I have known metaphysicians in heated moments fling the terms "abstractions" and "abstract thinking" at one another as a warning or a reproach. But to the mental tester there is no terror in the terms. Engaged as he is in the solution of severely practical problems, he sees no danger in abstract thinking. Indeed he suspects that all thinking worthy of the name is abstract thinking; and he knows that all measurements are measurements of abstractions. When he measures the length of a material object

he is well aware that the length is a pure abstraction —that it has no existence apart from the other dimensions, or indeed from the whole complex of qualities which in some mysterious way constitute the object. So when he measures intelligence he knows quite well that he is dealing with but one quality, or one group of qualities, of that highly complex thing called the mind, and not with a separate entity that exists in itself, by itself or for itself.

The argument goes on to urge that we cannot predict the efficiency of a man's mind in a large field of labour from observing its efficiency in a small field of labour; that when the milieu is changed the intelligence is changed; that, in fine, we have no right to say that he who is intelligent in little is intelligent also in much. A dullard at school may become a successful, nay, even a distinguished, man in the outside world; and the brilliant boy at college may in after life take his rightful place among the multitude of nobodies. To argue thus is to make the one the rule and the ninety-and-nine the exception. The truth is that the valuations of school and college are as a rule verified in after life. When they are falsified, as happens in rare instances, the explanation is simple: the pupils in question were misunderstood at school, where little account is taken of special interests and of peculiarities of temperament. There are certain elements of character which count for more in life than in school: initiative, for instance, and doggedness in the pursuit of one's purpose. And in this respect a mental test is no better than an examination; neither of them searches the heart of the child, nor does it

reveal the springs of conduct. The mental tests in present use do not pretend to go beyond the intellect. And yet the discerning examiner who, when using the Binet scale, comes into personal touch with the child cannot fail to learn something about the peculiarities of his character as well as about the workings of his mind.

The best refutation of the sceptic, however, lies in the realm of statistics. Professor Thorndike (who, by the way, is an uncompromising critic of the Spearman theory of the central factor and at the same time one of the most prolific inventors of mental tests) holds that our sole justification for making inferences from one kind of performance to another —from success in the examination room to success in life—lies in the correlation actually found between the two estimates of success. If there is a real correlation, positive or negative, we can prophesy with a degree of certainty that varies with the degree of correlation. When there is no correlation we cannot prophesy at all. This indeed is the only sound method. If a man wants to know what connection there is between the moon and the weather he may rely on his own observations, or ask an old sailor, or ask a philosopher, or ask his grandmother; but the only sure way of finding out is to study the meteorological records and to compare by statistical methods the changes in the weather with the changes in the moon. Thorndike does not believe in asking his grandmother. Neither did the Medical Department of the American Army. They were sceptical of the value of mental tests for soldiers until they were shown the high correlation between the results of a brief test in an examination hall and the results

of months of soldiering in the training camp and the field of battle. Then they were convinced. And when nearly two million men were so tested and the evaluations so confirmed, the whole nation was convinced as well.

Nobody claims that predictions based on mental tests are absolute and infallible, for it is realised that certain important factors are left out of the reckoning; but it is demonstrable that mental tests have greater predictive value than either examinations, or records of school progress, or the personal opinions of teachers and guardians. It is demonstrable that a child's performance on the miniature stage prepared for him by the mental tester is, generally speaking, an earnest of his performance in the larger theatre of life. What point therefore is there in discussing on *a priori* principles whether intelligence can be torn from its surroundings and still live—whether intelligence operating in one sphere remains the same if it is transferred to another sphere? Theories, however self-consistent they may be, must ultimately yield to the logic of fact. A disciple of Zeno's was once trying to convince Diogenes the Cynic of the impossibility of motion. Diogenes answered his argument by moving away.

Far too much store has been set on the argument that in comparing the intelligence of the man of 40 with that of the youth of 16 we have to place both subjects in a common intellectual world, which must perforce be the world of the youth, and in so doing we merge the man into a strange and unfamiliar universe. But why strange and unfamiliar? It is not a cloud-land remote from his own universe, but is itself a part of his own universe. It is the

common workaday world with which high and low, old and young, come into daily contact—the world where cakes and ale abound and where ginger is hot in the mouth. To take at random one of my own absurdities (the type of test at which the criticism has mainly been levelled)—the one about the old gentleman who said that he had always noticed that if he did not die in March he was all right for the rest of the year—can it be seriously contended that the elements of this problem lie within the lad's universe and outside the man's? If in handicapping the man we rob him of his superior strength, how comes it that in imposing a similar handicap upon the youth no such robbing takes place? Plunge the youth of 16 into the universe of the child of 6, and the child will be beaten at all points. So whatever theory of intelligence we embrace we must account for the fact that there is something more than the mere accumulation of knowledge—something which grows in the mind of the child and does not grow in the mind of the adult.

Those who feel easier when they can translate things mental into things physical may find in the fact that the brain reaches its maximum weight at 15 a comforting explanation of the theory I have tried to expand. But they will also have to face the disconcerting fact that by the middle of the fifth year the brain has already reached over 90 per cent. of its final weight.[1] It seems impossible therefore to establish a complete parallelism between increase of brain-weight and increase of intelligence.

Though we accept with equanimity the limita-

[1] I take these facts from Woodrow's *Brightness and Dullness in Children*, pp. 66–80.

tions of our bodies, the very notion of limitations in our minds rouses our hostility and sets our defences bristling. We love to think of the mind as a system of muscles that grow progressively and indefinitely bigger and stronger as we exercise them, or as a sword that gets ever brighter and keener as we use it. And if the doctrine of intelligence discredits this optimistic view by shutting off the vista of progress in the direction of intelligence, it retrieves the balance by opening out vistas in other directions. It is only after intelligence has ceased to grow, that the mind's most treasured lore begins to grow. There is no delusion in thinking that it is after 16 that we have acquired the bulk of the knowledge and culture and taste which we most highly value and most ardently cherish. And this is because as children we had to make our intellectual conquests by means of an intelligence that was imperfect and immature.

And now having said what I believe to be the broad truth of the matter, let me make a slight reservation. In climbing a very high mountain we begin by counting our progress in miles, but in course of time we come to count it in yards or even feet and inches; and the climbing which at first is pleasant and exhilarating at last becomes painful and exhausting; and few there are that reach the rarified air of the higher slopes. So it may be with our intelligence. Our measurements refer to the many; and the many will climb only when it is easy to climb: they will traverse the miles mapped out by heredity, but will not forge ahead among the feet and the inches. Dr. Maxwell Garnett in his monumental book on Education and World Citizenship gives cogent reasons for thinking that

voluntary attention—the power to attend to the uninteresting by sheer strength of will—is of the very essence of intelligence, and that voluntary attention can be cultivated. It may be so. It may be that the man who daily disciplines his thought can in this matter " burst his birth's invidious bar." But if so it means for him the girt loins, the trimmed lamp, and the watchful eye. And he has to be content with progress measured in inches, not in miles. It is not difficult to believe this doctrine of the trainability of attention. If it has never been proved, it at least has never been disproved.

CHAPTER XVIII

CORRELATION

It is not easy to follow the arguments used by experimental psychologists unless one has a fairly clear idea of what is meant by correlation. As used in statistical work correlation has nothing to do with Herbart, nothing to do with the snowdrop. It is a new—comparatively new—mathematical device for indicating the nexus between two measurable quantities: it tells how far we may expect one of them to vary concomitantly with the other. Although the question was first broached by Bravais nearly 80 years ago, it was not till Sir Francis Galton published in 1888 his paper on Correlations and their Measurement that the subject was seen to be of great scientific importance. Since that day the theory of correlation has been developed by Professor Karl Pearson and others into a means of mathematical analysis which has proved an inestimable boon to the psychologist. An attempt is made in this chapter and the next to explain in the simplest possible way the significance and use of correlation formulæ, and to give the reader sufficient knowledge to calculate correlations for himself and to interpret the results obtained by others. I give him an informal introduction to the subject: to know it well he must cultivate its acquaintance for himself. The chapter presupposes no special mathematical know-

ledge on the part of the reader—nor, for the matter of that, on the part of the author.

Let us begin by refusing to be overawed by a formula. Professor De Morgan in discussing the formula for the number of cannon balls in a pyramid of given base used to remind his hearers that the formula could tell them no more than could be discovered by counting the balls. Counting is the logical ground of all mathematics, including correlations. We will start therefore from the counting end.

Let us suppose that a teacher has a class of 15 boys (*a*, *b*, *c*, etc.), whom he examines in eight subjects (P, Q, R, etc.). He arranges the marks in columns as shown in Table III. It must be under-

TABLE III

Pupil.	P.	Q.	R.	S.	T.	V.	X.	Z.
a .	28	28	2	84	21	13	12	25
b .	25	25	5	75	20	28	12	27
c .	22	22	8	66	19	22	12	30
d .	20	20	10	60	18	5	12	23
e .	18	18	12	54	17	18	12	15
f .	17	17	13	51	16	14	12	13
g .	16	16	14	48	16	2	12	20
h .	15	15	15	45	15	8	12	9
i .	14	14	16	42	14	15	12	17
j .	13	13	17	39	14	10	12	19
k .	12	12	18	36	13	16	12	1
l .	10	10	20	30	12	25	12	18
m .	8	8	22	24	11	17	12	10
n .	5	5	25	15	10	12	12	6
o .	2	2	28	6	9	20	12	7
	—	—	—	—	—	—	—	—
Average	15	15	15	45	15	15	12	16

stood that the marks here given are wholly imaginary and are not at all likely to be obtained at a real examination: they are in fact specially chosen to illustrate certain principles.

Columns P and Q are identical. However well, or ill, a pupil does in one of these subjects, he does equally well, or ill, in the other. The correlation between them is positive and perfect, and is expressed mathematically as a correlation of 1. The usual method of representing marks of this kind geometrically is by means of a scatter diagram as in Fig. 19. Each point or cross represents a boy's position in subject P (horizontal axis) and in subject Q (vertical axis). This is the ordinary way of presenting the data; but at the suggestion of Dr. William Garnett I have adopted throughout this chapter a different mode of presentation: I have changed the origin of co-ordination from the point of intersection of the zero lines to the point of intersection of the mean lines. The mean lines now become the two axes, and all the measures are regarded as deviations. As the mean mark in column P is 15, *a*'s mark deviates from the mean by + 13, *b*'s mark by + 10, and so on until *o*'s mark is reached, which deviates by — 13. This new way of treating the measures is illustrated in Fig. 20, which represents the same data as Fig. 19, the only difference being that the origin is in a different place. The chief advantage of the change of origin is that it brings out the important fact that correlation depends wholly on the range of variation and in no way on the zero of measurement, which may be, and often is, a purely arbitrary datum. To quote from a letter of Dr. Garnett's, "The correlation between the heights

of mountains and the rainfall in the district is the same whether we measure the heights from Ordnance datum or from any other fixed level. As the range is the important thing it is reasonable that the

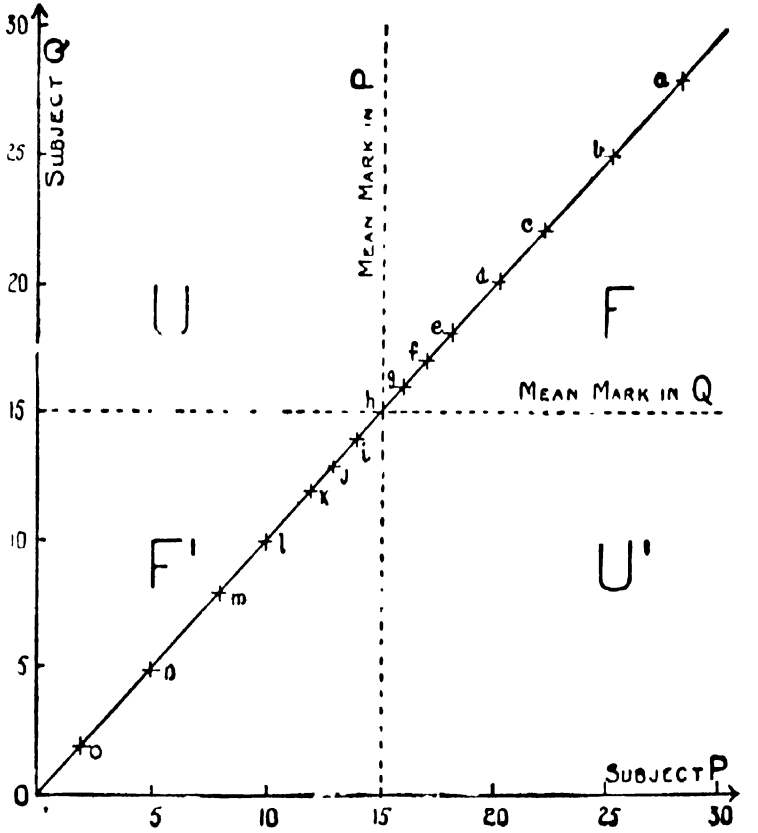


FIG: 19

centre of gravity of the range should be the origin The range then spreads more or less symmetrically, according to the distribution, both to the positive and negative sides."

Looking at Fig. 19 or Fig. 20, we see that the points representing the positions of the scholars lie in a straight line. Rarely, however, do the points all

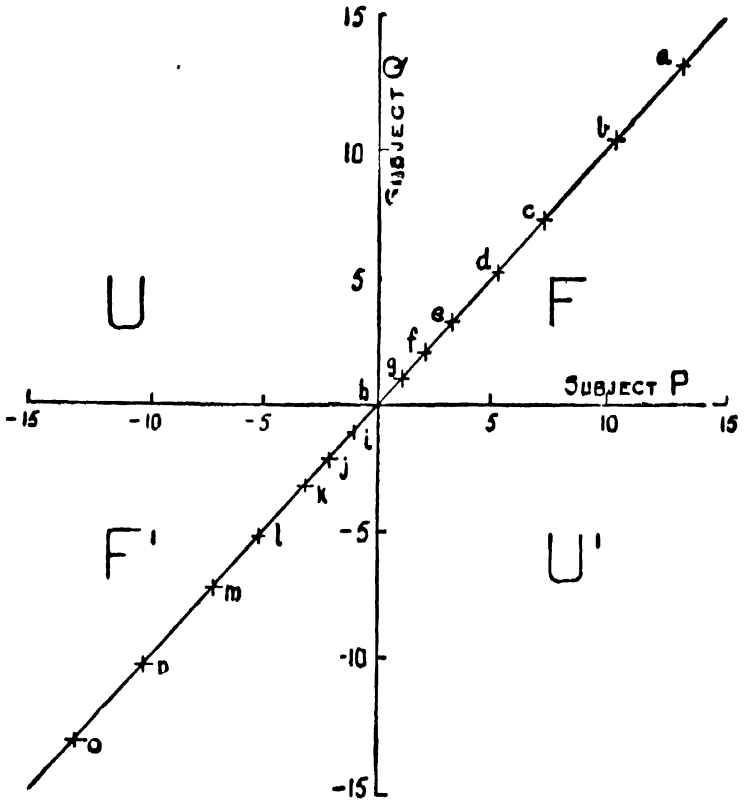


FIG. 20

lie in a straight line, though in most cases it is possible to find a line which shows the general trend—a best-fitting line which it is customary to call *the line of*

regression, from the fact that Galton in his writings on heredity used a geometrical device of this kind to show that the mean heights of offspring tended to "regress back towards the mean of the race." It

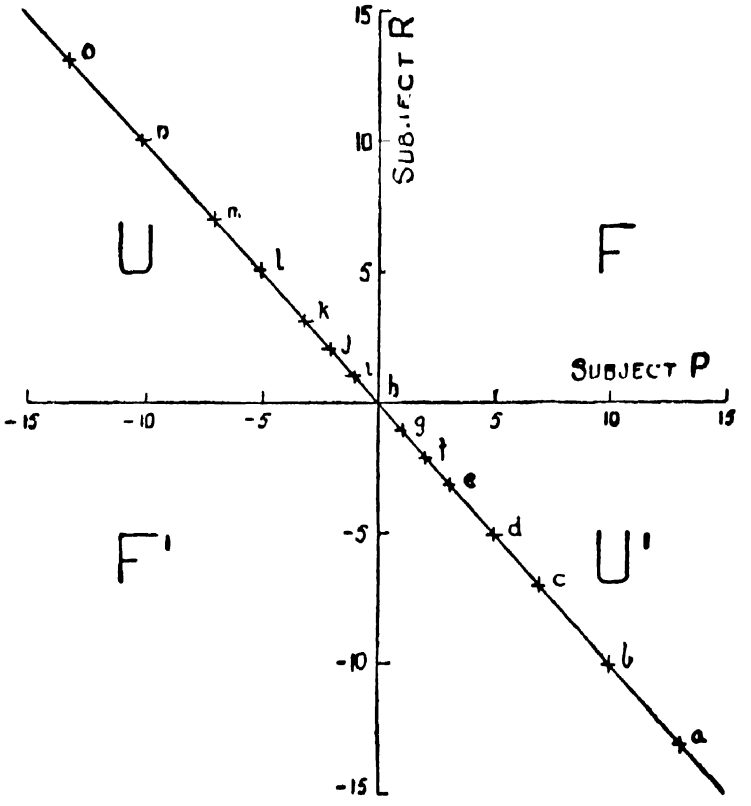


FIG: 21

will be observed that the line of regression here makes an angle of 45° with the axes and that the diagram is divided up into four quadrants F, F¹, U, and U¹.

Now compare with column P column R, which is column P turned upside down. The relationship

is perfect of its kind, but negative; a boy who does well in one of these subjects does badly in the other, and *vice versa*. Plot the measures as before (Fig. 21) and note that the line of regression is straight, that it makes an angle of 45° with the axes, and that all the points lie in the quadrants U and U^1. The correlation is — 1.

Now look at column S. Here the marks allotted to each pupil are three times as great as in column P. This may be due to many causes; one of the papers was three times as long as the other, or three times as hard relative to the standard of knowledge of the class, or one of the examiners was three times as liberal as the other. Whatever the cause, it is clear that there is a constant relation between the performances of the several pupils in the two subjects, and that the correlation is perfect, or unity. Taking the marks as they stand, let us see how they appear in a diagram (Fig. 22). The first thing that strikes us is that the line of regression no longer bisects the angle between the axes; it no longer slopes at an angle of 45°. But this line, the 45° line, is the line of perfect correlation, and any departure from that angle means a falling short of perfection. Indeed the angle which the line of regression makes with the nearer axis should represent that important ratio known as the coefficient of correlation, or *r*. If the angle is 45° the tangent is 1, and the coefficient 1; if the angle is less than 45° its tangent is less than unity, and the coefficient is less than unity. The tangent of the angle which the line of regression in Fig. 22 makes with the vertical axis is manifestly $\frac{1}{3}$, whereas theoretically it should be 1. If therefore the diagrammatic method is to be pursued some adjust-

ment of the measures is necessary in order to secure the right slope in the line of regression. This can

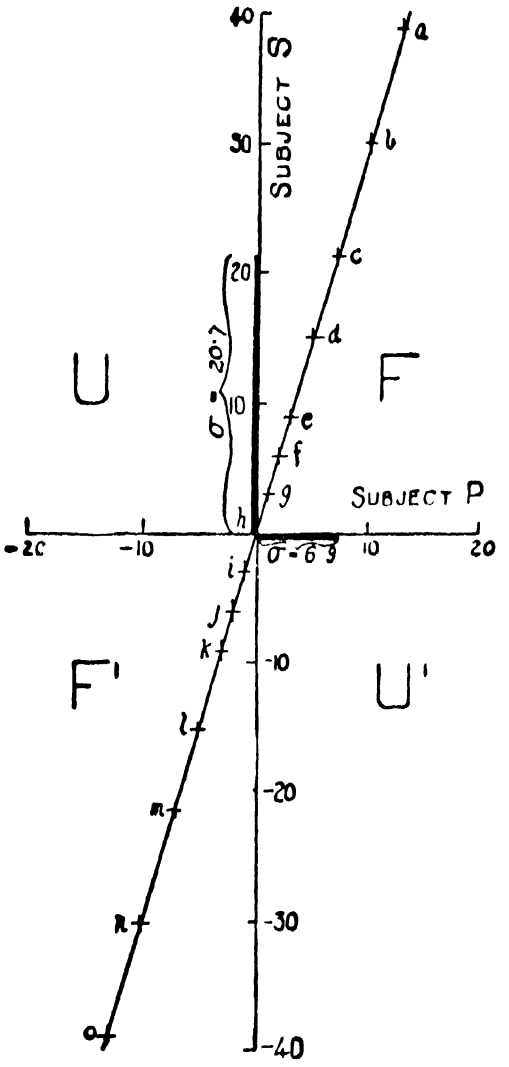


FIG: 22

obviously be secured by either dividing all the marks in S by 3, or by multiplying all the marks in P by 3.

When this is done the diagram becomes identical with Fig. 20.

It follows that the slope of the line of regression is not an index of the degree of correlation unless some adjustments are made among the measurements. Let us examine the nature of those adjustments. Looking at the diagram (Fig. 22), we see that if, in order to secure the 45° line of regression, we stretch the horizontal axis to three times its size, enlarging all its parts proportionally (or if we shorten the vertical axis to one-third of its size, reducing all its parts proportionally), we make the means of the two series of marks the same and we make the mode and degree of dispersion of the two series the same. Are both conditions essential? If, as was pointed out above, the position of the zero of measurement does not matter, it is clear that the average does not matter either. Suppose, for instance, that we added a certain number of marks, say ten, to each of the items in column P, Table III; we raise the average of the series by ten, but we in no way change the deviations. If the marks in column P, as it now stands, and the marks increased severally by ten, were plotted in a diagram, that diagram would differ in no respect from Fig. 20. The absolute value of the average does not matter: it is the dispersion alone that is significant.

The measure of dispersion that is nowadays used in calculating the coefficient of correlation is, for reasons that need not detain us here, the standard deviation, or σ. The simplest measure of dispersion is the mean deviation, which is merely the average of the deviations that appear in the scatter

diagrams in this book. To find the standard deviation, the several deviations must be squared before finding the mean; and the square root must be taken after finding the mean. The standard deviation of the marks in P is 6·9, of the marks in S 20·7—three times as great. The essential condition, therefore, of which we are in search is that the standard deviation of the two series should be the same. This can be secured by taking the standard deviation of each series as the unit of measurement of that series—a device of which more will be said later.

This principle may be further illustrated by reference to column T in Table III. Here the boys come out in the same order of merit as in P, but the marks are differently distributed. In T they cluster more closely round the mean. The measures (graphed in Fig. 23) all appear in the quadrants F and F^1, and although they do not all lie exactly in the line of regression they never diverge widely from it. The correlation is manifestly almost as perfect as in Fig. 20: it is in fact nearly 1. There is however the same peculiarity in the slope of the line of regression as was noticed in the last example; and it is manifestly due to the same cause—the different degrees of dispersion in the two series. The standard deviation of the marks in P is 6·9, and of the marks in T 3·5. One is in fact nearly twice as large as the other. If therefore we use the standard deviation of the group as the unit of measurement for that group, the unit for P covers twice as many of the teacher's marks as the unit for T. To make the distributions equal and the marks comparable we must in plotting the measures give each mark in T twice as much space

as a mark in P. This is done in Fig. 24. What we have done is to stretch out the more compressed series so as to make it as widely distributed as the other. And again, as in the previous instance, we alter the slope of the line of regression. Then we got it perfect; now, nearly perfect.

Galton used a diagrammatic scheme of this kind;

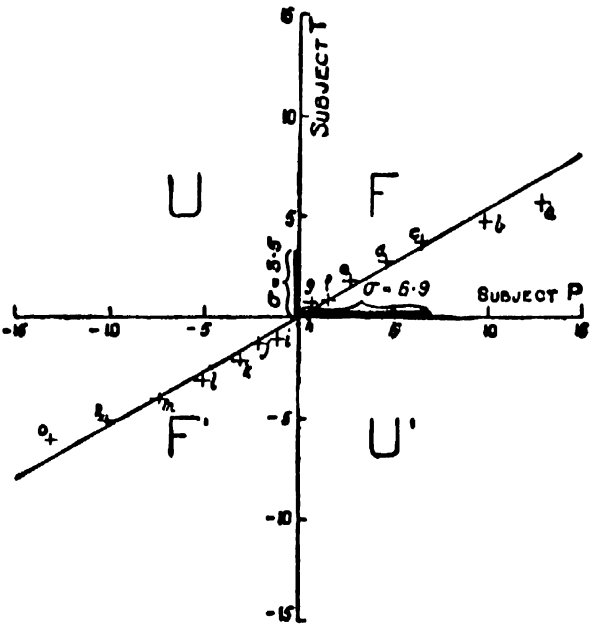


FIG. 23

but his method of adjusting the distributions of the two series was to make the interquartile ranges[1]

[1] The interquartile range includes the middle half of the number of cases. If a teacher arranges a set of examination papers in order of merit and divides them into two equal groups, the mark obtained by the middle paper of the lower group is the lower quartile, and the mark obtained by the middle paper of the upper group is the upper quartile. The difference between these two marks is the interquartile range.

equal; not, as we have done, to make the standard deviations equal. But it amounts to the same thing. For in a normally distributed series the standard deviation bears a constant ratio to the interquartile range. To make one pair equal is to make the other pair equal. And when this is done the tangent of the angle which the line makes with the nearer axis becomes the coefficient of correlation, or *r*.

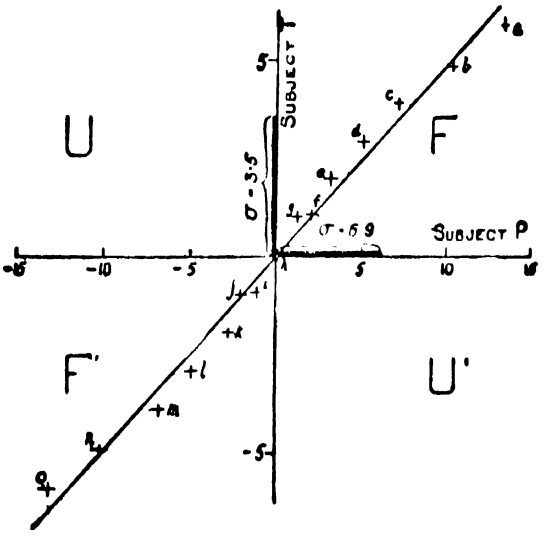


FIG 24

We seem justified in concluding that if the points representing the positions of the candidates in the two subjects of examination lie in a straight line, and that line does not bisect the angle between the axes, then the dispersion of the marks in one subject differs from the dispersion of marks in the other subject, and when this difference is abolished the line of regression immediately falls into the 45°

position. If the reader will try to plot a linear series of points which shall not coincide with the line of perfect correlation and at the same time shall represent marks of equal dispersion, he will realise the impossibility of the task.

But although we cannot find one such row of points which fail to coincide with the line of perfect correlation (positive or negative), we can find *two* such lines. Look at Fig. 25, for instance, where I have plotted 29 points with the following co-ordinates, expressed, as in the other diagrams, in terms of deviations from the means: (27, 9), (21,7), (18, 6), (12, 4), (9, 27), (9, 3), (7, 21), (6, 18), (6, 2), (4, 12), (3, 9), (3, 1), (2, 6), (1, 3), (0, 0), (−1, −3), (−2, −6), (−3, −1), (−3, −9), (−4, −12), (−6, −2), (−6, −18), (−7, −21), (−9, −3), (−9, −27), (−12, −4), (−18, −6), (−21, −7), (−27, −9)

The standard deviation is the same in both series (11·6). Here we have two best-fitting lines—two lines of regression—each diverging equally from the line of perfect correlation, each making with the nearer axis an angle whose tangent is one-third. The coefficient of correlation is accordingly one-third or ·3̇. This diagram may be taken as an ideal type of all scatter diagrams—ideal because it is a purely hypothetical example devised to make manifest certain features that are obscured by the helter-skelter distribution of points in the diagram arrived at in experimental work. There are always, as a matter of fact, two lines of regression, except when the correlation is perfect; and when the correlation is perfect we may suppose the two lines to have coalesced. The usual method of finding

the regression lines in a scatter diagram is to plot a much larger number of points than I have placed in any of these diagrams, and then to divide the surface into a number of horizontal strips of equal and uniform width, a width representing, say, 2, 5, or 10 units. The mean position of the points in each

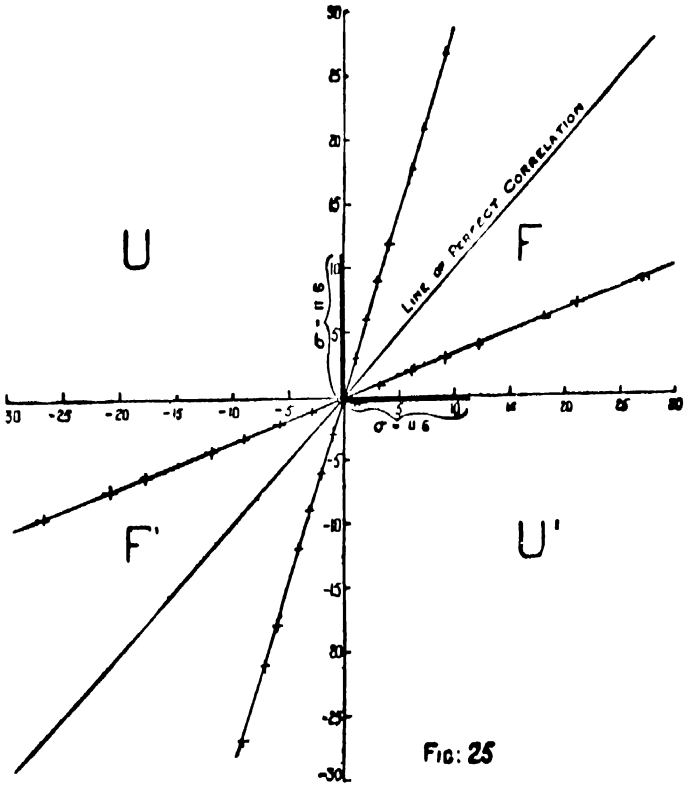


FIG. 25

strip is then found, and the direction of these mean points gives one line of regression. The other line of regression is found in a similar way by taking vertical strips instead of horizontal strips. Of the two lines of regression, one shows the regression of y on x and indicates the most probable value of y

for any given value of x; the other shows the regression of x on y and indicates the most probable value of x for any given value of y.

Let us now consider the relation between column P and column V. There is no perceptible connection between them except that the same numbers appear in both. And this is not surprising; for the

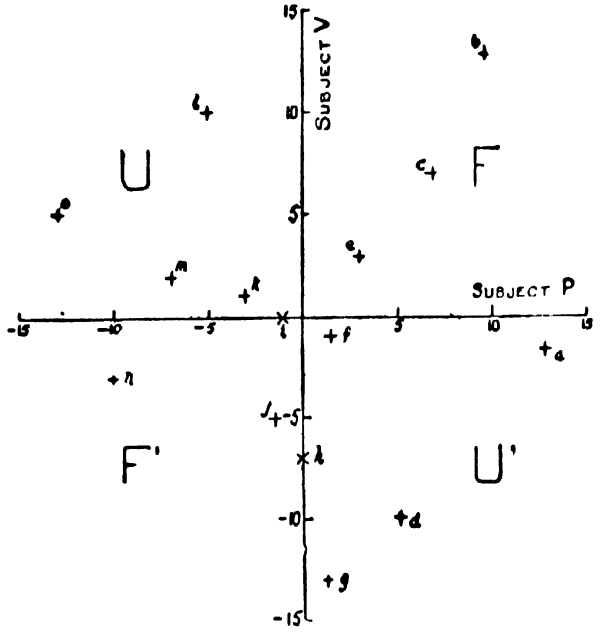


FIG 26

order in column V was determined by writing the numbers in column P on dried beans, shaking the beans up in a bag, and taking them out at random. The effect of plotting the two series is seen in Fig. 26. It will be noticed that the points are widely scattered and are almost equally distributed among the four quadrants. Moreover, it seems impossible to find by inspection a best-fitting line. We may

assume, therefore, that the correlation is low. As a matter of fact it is nearly zero. It must not be supposed, however, that the points must necessarily be widely scattered when the correlation is nought; for the correlation between column P and column X is obviously nought; there are no changes of value in the items of X to correspond with the changes of value in the items of P. And yet the graph

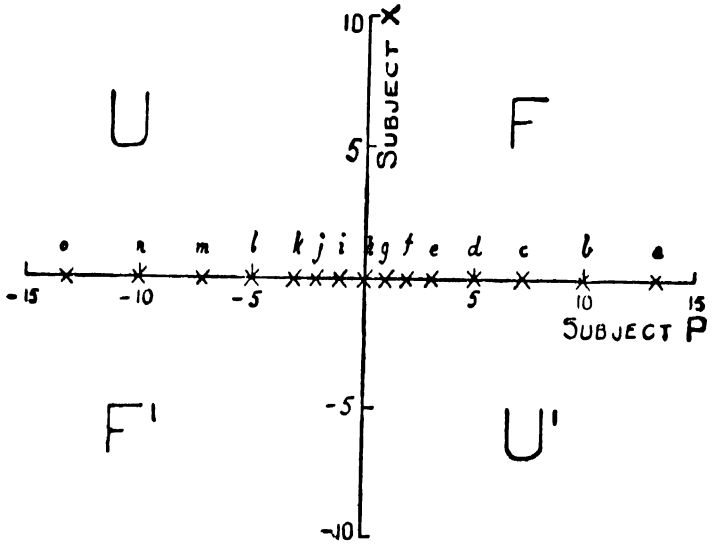


FIG: 27

(Fig. 27) shows that all the points lie in a straight line, that line being in fact one of the axes. To equalise the distributions of the two series is here impossible, for in one of the series there are no variations. In the previous diagram, however (Fig. 26), there is no disparity in the degree of dispersion, for the standard deviation in each series is the same (6·9). If we had a sufficient number of points to adopt the

method of horizontal and vertical strips as described above, we should find in this instance the lines of regression tending to coincide with the two axes. As the line bisecting the quadrants F and F^1 is the line of perfect positive correlation, and the line at right angles to it and bisecting the quadrants U and U^1 the line of perfect negative correlation, so are the two axes the lines of regression which indicate the absence of correlation. And the degree to which regression lines approach these limiting positions serves as a clue to the relationship between the variates compared.

So far we have discovered that when the measures of the correlated series are set forth in a graph, those that appear in two of the quarters are favourable to a positive correlation, and those that appear in the other two are favourable to a negative correlation. When they are almost equally distributed among the four quarters the correlation approximates zero. When the distributions of the two series are adjusted by using the standard deviation of each as its unit of measurement, the slope of the regression lines affords an index of the degree of correlation between the series.

I will now give the Bravais-Pearson product-moment formula for the coefficient of correlation (r), and see how it fits in with our findings :

$$r = \frac{\Sigma xy}{N\sigma_1\sigma_2}.$$

In this formula x and y stand for the deviations of each of the measures from the mean value of its series, σ_1 for the standard deviation of the first series, σ_2 for the standard deviation of the second series, Σ

for summation, and N for the number of things or persons measured. The significance of the formula can best be studied by applying it to a specific instance. Let us apply it to find the correlation between the marks in column P and the marks in column Z in the example on p. 180. The various numerical steps set forth should be compared with the corresponding graph (Fig. 28). The average of series P is 15, and column x gives the deviations of each measure in the series from 15. Thus *a* deviates by 13, *b* by 10, etc. In the same way column y gives the deviations of Z from its average, 16. The product of x and y for each pupil is placed in the last column; and it is the algebraic sum of these products (642) that forms the numerator of the fraction in the formula. When a pupil's deviations from the averages are positive in both series (*a*'s deviations, for instance), or negative in both, like *m*'s, his position in the graph lies in one of the quadrants F and F^1, and the product of his deviations, being positive, helps to swell the positive element in the numerator and to increase the degree of correlation. When on the other hand a pupil's variations are of different signs, such as those of *f* and *l*, his diagrammatic quadrant is U or U^1, and the product of his deviations becomes negative, helping to swell the negative element in the numerator and to reduce the degree of correlation. Moreover, the nearer the diagonals a point lies, the more does the product of its deviations contribute towards deciding the nature and magnitude of the numerator in comparison with the contribution which the squares of these deviations make towards the magnitude of the denominator.

EXAMPLE OF CORRELATION (PRODUCT-MOMENT FORMULA)

Pupils	P.	Z.	x.	y.	x^2.	y^2.	xy.
a	28	25	13	9	169	81	117
b	25	27	10	11	100	121	110
c	22	30	7	14	49	196	98
d	20	23	5	7	25	49	35
e	18	15	3	−1	9	1	−3
f	17	13	2	−3	4	9	−6
g	16	20	1	4	1	16	4
h	15	9	0	−7	0	49	0
i	14	17	−1	1	1	1	−1
j	13	19	−2	3	4	9	−6
k	12	1	−3	−15	9	225	45
l	10	18	−5	2	25	4	−10
m	8	10	−7	−6	49	36	42
n	5	6	−10	−10	100	100	100
o	2	7	−13	−9	169	81	117
	15	16			714	978	642

$$\sigma_1 = \sqrt{\frac{714}{15}} = 6{\cdot}9 \qquad \sigma_2 = \sqrt{\frac{978}{15}} = 8{\cdot}07$$

$$r = \frac{\Sigma xy}{N\sigma_1\sigma_2} = \frac{642}{15 \times 6{\cdot}9 \times 8{\cdot}07} = \frac{642}{835} = {\cdot}77$$

The labour of evaluating σ_1 and σ_2 is saved by using a formula which is equivalent to the original one.

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}} = \frac{642}{\sqrt{714 \times 978}} = {\cdot}77$$

The quadrants of the graph which I have named F and F¹ are favourable to positive correlation; the quadrants which I have named U and U¹ are unfavourable to positive correlation. The numerator Σxy expresses the balance of the distribution

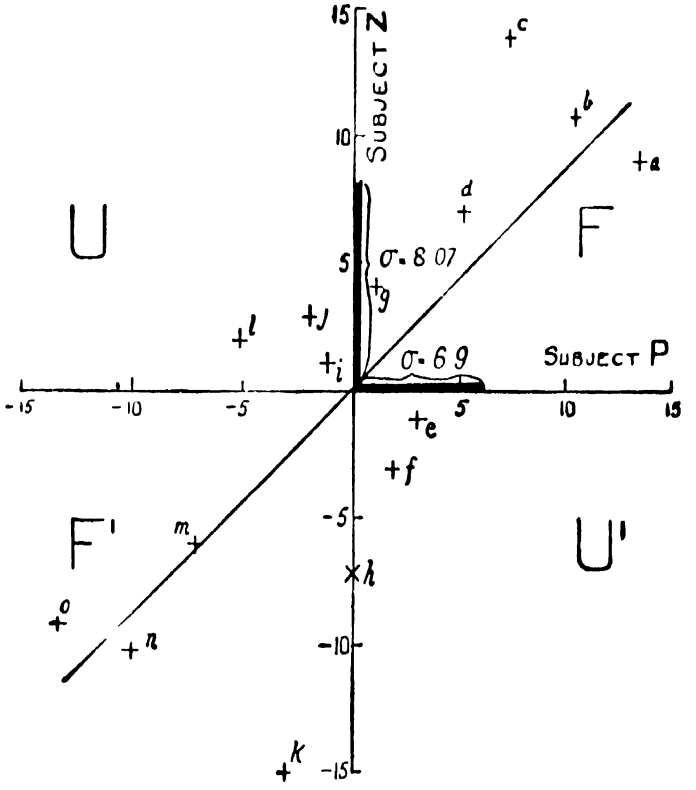


FIG: 28

of points among the favourable and unfavourable quarters. The denominator $N\sigma_1\sigma_2$ secures that adjustment of distribution of measures in each series which I have shown as necessary to determine the proper slope of the lines of regression.

For calculating the coefficient of correlation there are two labour-saving devices with which the reader should make himself familiar. The first is the use of the formula $r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \times \Sigma y^2}}$ instead of $r = \frac{\Sigma xy}{N\sigma_1\sigma_2}$.

To appreciate the advantage of the former the reader should apply both to the data furnished in the example on p. 180. That the two formulæ are equivalent may be shown thus:

$$N\sigma_1\sigma_2 = N\sqrt{\frac{\Sigma x^2}{N}} \times \sqrt{\frac{\Sigma y^2}{N}} = N\sqrt{\frac{\Sigma x^2 \times \Sigma y^2}{N^2}}$$

$$= \sqrt{\Sigma x^2 \times \Sigma y^2}.$$

The second labour-saving device bears on the setting forth of the data. It will be found in practice that the mean of the measures will rarely work out exactly: there will be a fraction, which, by the way, it is not necessary to carry beyond the first place of decimals. This complicates the process of computing the deviations and the squares of the deviations. The complication may be avoided by taking as a provisional mean the nearest whole number, and restoring the balance afterwards. Let us suppose, for instance, that in column P (p. 180), the actual mean is not 15 but 15·3, and in column Z not 16 but 15·8. We could then complete the calculations thus:

	P	Z
Means . .	15·3	15·8
Provisional means	15	16
Difference . d_1 =	·3	d_2 = −·2

	x^2	y^2	xy
	$\Sigma = 714$	978	642
Subtract	$Nd_1^2 = 1{\cdot}35$	$Nd_2^2 = 0{\cdot}6$	$Nd_1d_2 = -0{\cdot}9$
	712·65	977·4	642·9

$$r = \frac{\Sigma xy}{\sqrt{\Sigma x^2 \Sigma y^2}} = \frac{642{\cdot}9}{\sqrt{712{\cdot}7 \times 977{\cdot}4}} = {\cdot}77.$$

It will be observed that if we had merely taken the round numbers for the means and made no compensation at all it would not have made a difference of ·01 in the result.

Having taken the trouble to compute the correlation between the marks obtained by 15 pupils in two hypothetical branches of study P and Z, I will now point out that the result arrived at is worthless —or almost worthless. It is worthless because the number of pupils is so small. For it must be remembered that when we are in search of a correlation coefficient we are mainly concerned, not with discovering the correlation between the achievements of a particular group of pupils, but with discovering the correlation between two branches of study or two types of ability; and unless the coefficient obtained holds good (within a reasonable margin of error) for other groups of pupils tested at other times—unless it has general applicability as a trustworthy measure of the concomitant variation of the factor tested, it has no value whatever as a basis of inference. There is no known method of finding with absolute certainty the correlation between two variables, if only one sample is tested. The best method is the product-moment method which I

have just explained; but this method is based on certain assumptions which are at best only approximately true. It assumes, for instance, that each of the factors to be compared and correlated is normally distributed—that it follows pretty closely the normal curve of error—and further that the particular group measured is a good sample, a representative sample, a sample which exhibits the general features of the complete group. There is little danger in assuming that any type of ability which the teacher measures conforms to the normal curve (nature has so frequently been found to distribute her gifts in this particular way that the assumption is amply justified); but there is much danger in assuming that the few cases tested fully represent the whole class of things to which they belong. There is one thing that we may safely say, and that is that the larger the number of cases tested, provided of course they are taken at random and not selected, the more trustworthy will be the results. When the number of cases is small the special features of the normal distribution curve are unlikely to appear, and standard deviation is robbed of much of its significance. When the number of cases is large enough to display, roughly at least, the characteristics of normal distribution, the mere statement of the mean and of the standard deviation reveals the general shape of the curve, and inferences can legitimately be drawn from that statement. The standard deviation can then too be advantageously used as a unit of measurement for the group. It is generally agreed that when the number of cases is less than 30 even the product-moment formula will fail to yield a trustworthy coefficient of correlation.

Besides the product-moment formula for correlation, several other formulæ have been suggested which aim at giving with as little calculation as possible an approximate measure of the correlation. Such methods are of particular value when N is less than 30, for the approximate results so obtained are likely to be as trustworthy as those obtained by the more laborious Bravais-Pearson method. One of these simpler methods is the method of ranks. It can be shown that if instead of giving absolute scores for the performances of the individuals of a group we merely rank them in order of merit, the product-moment formula for r reduces to $r = 1 - \frac{\Sigma d^2}{\frac{1}{6}N(N^2-1)}$, where d means the difference in the rank of an individual in the two series of performances. But the formula may be arrived at independently; and by attempting to arrive at it by a second route I hope to make its significance clearer.

It will be seen that—

$$1^2 + 3^2 = \tfrac{1}{6} \times 4 \times (4^2 - 1)$$
$$2^2 + 4^2 = \tfrac{1}{6} \times 5 \times (5^2 - 1)$$
$$1^2 + 3^2 + 5^2 = \tfrac{1}{6} \times 6 \times (6^2 - 1)$$
$$2^2 + 4^2 + 6^2 = \tfrac{1}{6} \times 7 \times (7^2 - 1)$$

and, generally, the sum of the squares of consecutive even numbers (beginning with 2) or consecutive odd numbers (beginning with 1) whose last term is $N - 1$ is $\frac{1}{6}N(N^2 - 1)$.

Let us now suppose 7 pupils (*a*, *b*, *c*, etc.) to be examined in two school subjects P and Q and arranged in order of merit as in Table IV. If the order is the

TABLE IV

Pupil.	Rank in P.	Rank in Q.	d^2.	Rank in R.	Rank in S.	d^2
a .	1	7	6^2	1	8	7^2
b .	2	6	4^2	2	7	5^2
c .	3	5	2^2	3	6	3^2
d .	4	4	0^2	4	5	1^2
e .	5	3	2^2	5	4	1^2
f .	6	2	4^2	6	3	3^2
g .	7	1	6^2	7	2	5^2
h .	–	–	–	8	1	7^2

same in both subjects the difference in the ranks of each pupil is nought. $d = 0$, $d^2 = 0$ and $\Sigma d^2 = 0$. When, however, the order in one subject is exactly the reverse of the order in the other, Σd^2 is at its maximum. I have in Table IV made the order in Q the reverse of the order in P; and I have done the same for eight pupils in subjects R and S. By inspecting the columns under d^2 in the table, and comparing them with the series dealt with above, it will be seen that when Σd^2 is greatest it amounts to twice $\frac{1}{6}N(N^2 - 1)$ or to $\frac{1}{3}N(N^2 - 1)$. When the ranks are the same the correlation is 1 and $\Sigma d^2 = 0$; when the ranks are as different as they can be the correlation is -1 and $\Sigma d^2 = \frac{1}{3}N(N^2 - 1)$. What would Σd^2 be when the correlation is 0; that is, half way between 1 and -1? Presumably half $\frac{1}{3}N(N^2 - 1)$ or $\frac{1}{6}N(N^2 - 1)$. This conclusion is confirmed by the fact that $\frac{1}{6}N(N^2 - 1)$ is the value that the sum of the d^2's would have if the ranks were determined by taking numbered counters at random out of a

bag; or in other words if the position of the pupils were settled by chance alone.

Returning to the original formula

$$\rho = 1 - \frac{\Sigma d^2}{\frac{1}{6}N(N^2 - 1)}$$

we readily see that when $\Sigma d^2 = 0$, $\rho = 1$; when it equals $\frac{1}{3}N(N^2 - 1)$, $\rho = -1$; and when it equals $\frac{1}{6}N(N^2 - 1)$, $\rho = 0$. It thus validly applies to the two extremes and the middle, and may be presumed to afford a comparative measure of intermediate degrees of correlation. The formula is usually put in the neater form

$$\rho = 1 - \frac{6\Sigma d^2}{N(N^2 - 1)}.$$

In computing correlation by the method of ranks a difficulty arises when two or more pupils are bracketed. It is customary in this case to give each

EXAMPLE OF RANK CORRELATION

Pupils.	Rank in P	Rank in Q	d^2.	
a .	1	2	1	$\rho = 1 - \frac{6\Sigma d^2}{N(N^2-1)}$
b .	2	3	1	
c .	$3\frac{1}{2}$	6	$6\frac{1}{4}$	$= 1 - \frac{6 \times 91\frac{1}{2}}{10 \times 99}$
d .	$3\frac{1}{2}$	10	$42\frac{1}{4}$	
e .	5	1	16	
f .	6	4	4	$= 1 - \frac{549}{990}$
g .	7	9	4	
h .	8	6	4	$= 1 - .55$
i .	9	6	9	$= .45$
j .	10	8	4	
			$91\frac{1}{2}$	

pupil in the bracketed group the average rank of the group. In the hypothetical example which I have worked in the accompanying example by the method of ranks pupils *c* and *d* are bracketed together in subject P and pupils *c*, *h*, and *i* in subject Q.

The method of ranks or position was first used by Professor C. Spearman and has been advocated by him as an excellent method for dealing with psychical as distinct from physical measurements. He has also suggested a still simpler formula, called the foot-rule formula. It is :

$$R = 1 - \frac{\Sigma g}{\frac{1}{6}(N^2 - 1)}$$

where *g* means the gains in rank. In the example just worked the gains are made by *e*, *f*, *h*, *i*, and *j*, and they amount respectively to 4, 2, 2, 3, and 2; i.e. $\Sigma g = 13$. The correlation between P and Q works out by this formula thus :

$$R = 1 - \frac{6 \times 13}{100 - 1} = 1 - \frac{78}{99} = 1 - \cdot 79 = \cdot 21$$

It would have made no difference if we had taken the losses in rank instead of the gains; for what one gains others lose and the losses and gains are necessarily equal.

The method of ranks has been criticised on the ground that it postulates a mode of distribution among the variates which is never found in nature : it assumes that the differences between the ranks are all equal—that the step from the first to the second is exactly as large as the step from the second to the third, and so on all along the series. But we know that the actual distribution of traits in nature does not fit in with this schema—that the differences

then tend to be small towards the middle or mean of the group and large towards the extremes.

One reply to this criticism is that the method is generally applied when the number of cases is small, and it would be just as unwise to assume any other form of distribution as to assume the rectangular form. Another reply is that the method of rank is used when it is difficult if not impossible to assign absolute marks to performances, and the most one can do is to say that one performance is better or worse than another.

Statisticians, however, insist on making allowance for the difference in distribution assumed in computing ρ and r respectively, and they convert ρ into r by means of Pearson's formula: $r = 2 \sin\left(\frac{\pi}{6}\rho\right)$.

They save us the trouble of calculating from the formula by providing us with a table which gives the values of r that correspond with the various values of ρ. An examination of this table shows that r is always the larger of the two, except when the values are 1, 0, and -1, when they are identical. The difference, however, is very small: it never amounts to ·02. It is the largest when $\rho = ·57$, and the corresponding value of r is ·5881. To be so particular about the second place of decimal when the first place is singularly uncertain is a kind of spurious accuracy which is not uncommon in the realm of statistics. The prevailing fashion is, however, to convert ρ into r, and if the reader wishes to follow the fashion, he need not trouble about the formula, but simply add ·01 to the value of ρ when it falls between ·1 and ·95, and leave it unchanged when it

falls outside that range. In the example on p. 187 the conversion is effected by changing $\rho = \cdot 45$ into $r = \cdot 46$.

The foot-rule formula, however, is quite another matter. Professor Spearman never intended it to be used for precise purposes, but in order "to gain quickly an approximate valuation of r." Indeed the reader, if he will deal with R as I have dealt with ρ—if he will experiment with the reversing of ranks—will discover that for negative correlations R is valueless. He will find that when N is odd it is impossible to get a negative correlation greater than $-\cdot 5$. When N is even the maximum correlation obtainable diminishes as N increases. And the table for the conversion of R into r shows that there can be as much difference as ·23 in the corresponding values. It is to the credit of R that it tends to underestimate the value of the coefficient and thus leans to the side of caution.

This foot-rule method of Professor Spearman's has proved a great boon as a time-saver. It is true that it has been arrived at empirically and that it lacks the theoretic soundness of the product-moment method; but those who have used it widely and compared the results with those obtained by other methods have found that, with small groups at any rate, it "works" remarkably well. The fact that it is unsuitable for negative correlations need not disturb the teacher, for negative correlations—high ones at any rate—are never found among those mental abilities which he is concerned in measuring. Since, however, the difference between R and r is sometimes considerable, it is well to make the coefficient comparable with other coefficients by converting R into r from the formula

$$r = 2\cos\frac{\pi}{3}(1 - R) - 1,$$

or, more conveniently, from Table V.

TABLE V

The conversion of R into r $\left(r = 2\cos\frac{\pi}{3}(1 - R) - 1\right)$.[1]

R	r	R	r	R	r	R	r
·01	·018	·26	·429	·51	·742	·76	·937
·02	·036	·27	·444	·52	·753	·77	·942
·03	·054	·28	·458	·53	·763	·78	·947
·04	·071	·29	·472	·54	·772	·79	·952
·05	·089	·30	·486	·55	·782	·80	·956
·06	·107	·31	·500	·56	·791	·81	·961
·07	·124	·32	·514	·57	·801	·82	·965
·08	·141	·33	·528	·58	·810	·83	·968
·09	·158	·34	·541	·59	·818	·84	·972
·10	·176	·35	·554	·60	·827	·85	·975
·11	·192	·36	·567	·61	·836	·86	·979
·12	·209	·37	·580	·62	·844	·87	·981
·13	·226	·38	·593	·63	·852	·88	·984
·14	·242	·39	·606	·64	·860	·89	·987
·15	·259	·40	·618	·65	·867	·90	·989
·16	·275	·41	·630	·66	·875	·91	·991
·17	·291	·42	·642	·67	·882	·92	·993
·18	·307	·43	·654	·68	·889	·93	·995
·19	·323	·44	·666	·69	·896	·94	·996
·20	·338	·45	·677	·70	·902	·95	·997
·21	·354	·46	·689	·71	·908	·96	·998
·22	·369	·47	·700	·72	·915	·97	·999
·23	·384	·48	·711	·73	·921	·98	·999
·24	·399	·49	·721	·74	·926	·99	·999
·25	·414	·50	·732	·75	·932	1·00	1·000

[1] Quoted from the Appendix to Rugg's *Statistical Methods Applied to Education*, p. 402.

When we divide our pupils into two groups only (such as those who wear white collars and those who do not, those who can read and those who cannot, those who understand logarithms and those who do not, or those who pass a simple test and those who fail), and wish to find the connection between two qualities, the most convenient formula to use is that recommended by Mr. Burt in Appendix II of *Mental and Scholastic Tests.* It is called the coefficient of colligation (ω).

$$\omega = \frac{1 - \sqrt{\frac{U\,U^1}{F\,F^1}}}{1 + \sqrt{\frac{U\,U^1}{F\,F^1}}}$$

F here means the number of instances when both the positive qualities compared are present, F^1 the number of instances when both are absent, U and U^1 the number of instances in which one is present and the other absent. It will readily be seen that large F's are favourable to positive colligation, that large U's are favourable to negative colligation, and that the formula falls in roughly with the scheme of quartering in our scatter diagrams.

In the finding of correlations there are many pitfalls. We have to be careful in many ways. We have to be careful to take a random sample instead of a picked sample and a large number of cases instead of a small; we have to give no wider interpretation to the results than the data warrant; and, most important of all because it is so frequently overlooked, we must be careful to measure as accurately as the conditions will allow. For unless

our measurements are good and true no superstructure that we build upon them can be good and true. If our premises are false, our conclusions can be true only by accident. To calculate by laborious means the correlation between two series of marks arrived at by arbitrary or unscientific methods is pure waste of time; to hope to secure trustworthy results therefrom is as futile as to expect white from two blacks or right from two wrongs. No correction formula can neutralise the effect of careless testing and careless marking; and since the testing and the marking usually carried out in the schools fall short of that precision which science demands, the teacher who wishes to estimate correlations between various forms of ability has first to learn how to test.

When the number of persons tested is small, it does not much matter which formula is used for finding correlations: between these unveracious witnesses there is not much to choose. Yet, although the coefficients cannot justly testify to the exact degree of correlation, they can at least point out the existence of correlation; and they afford a rough index of the amount. And since ρ and R are much easier to calculate than r, with small numbers the rank method should be used in preference to the product-moment method. Using the data furnished in the example on p. 180 we find $r = \cdot 77$, $\rho = \cdot 80$ and $R = \cdot 52$. Converting all into r according to the transmutation formulæ, we find the three values to be severally ·77, ·81, and ·75. In experimenting with group tests on American lines I applied a set of eight tests to a class of 38 girls, and the following week applied another similar

set. The coefficients of correlation between the two results were calculated by the three methods and the following values obtained : $r = \cdot 74$, $\rho = \cdot 74$, and $R = \cdot 53$. Converting all into r we get ·74, ·75 and ·76. There are, therefore, good grounds for thinking that when the numbers are small there is nothing to be gained by choosing the most laborious mode of calculation.

I have assumed all along that the regression is linear—that the line which indicates the general trend of the dots in the scatter diagram is a straight line. But the line of regression is not always a straight line : it is sometimes a curve. There is, for instance, a close relationship between the volume of a gas and the pressure to which it is subjected—a relationship expressed in Boyle's Law, which says that the temperature remaining constant the volume of a gas varies inversely as the pressure. If for a pressure of one unit the volume is 60 units, the volumes for 1, 2, 3, 4, and 5 units of pressure would be 60, 30, 20, 15, and 12 respectively. When these volumes are plotted the points will be found to lie in the path of a curve. The line of regression is not straight, but hyperbolic. For cases of this kind the product-moment formula is not applicable, and a distinct formula has to be used which gives the correlation ratio, or η. Since, however, in those psychological and educational measurements with which we are here concerned the regression is nearly always linear, the reader would be well advised to devote but little attention to the correlation ratio in his early studies. It suffices at that stage to be aware of its existence.

There is a mathematical test for linearity—a test

which involves a fair amount of calculation. The reader will, however, find it more convenient (and perhaps more convincing) if he makes a scatter diagram of his data and judges for himself whether or not the general trend approximates to a straight line.

CHAPTER XIX

PROBABLE ERROR

MEN of letters have looked long askance at the word "reliable"; they doubted its respectability.[1] Statisticians, untroubled by such doubts, have taken the word to their bosoms, given it definite meaning, and made abundant use of it. A reliable test is one that yields the same results again and again. If to-day we measure with a rule the lengths of the separate books on a book-shelf, and to-morrow measure them again, we shall find that the two results tally. If they do not tally we conclude that we have measured carelessly. There was no lack of reliability in the test, but only in the testing. When, however, we try to measure a mental quality we get far less uniformity in our findings. In applying the

[1] My colleague, Dr. J. H. Jagger, has reminded me that "reliable" is of more ancient and more respectable pedigree than its synonym "trustworthy." The instances cited in the *New English Dictionary* show that "trustworthy" appeared in the world of letters 260 years after its rival and is in comparison a parvenu and an upstart. The objection to the structure of the word "reliable" is of doubtful validity. If "reliable" should be rely-on-able, thus should "laughable" be laugh-at-able, "dependable" depend-on-able, and "available" avail-oneself-of-able. The critics as a final resort point to the kind of people who use the word: they are the kind who misuse "individual" and "aggravate," and are generally lacking in a sense of linguistic propriety. To this argument the reply of the statistician can be summed up quite briefly: it is "Tut! tut!"

same test to the same pupils under what seem to be the same conditions, rarely do we secure the same results. Nor indeed do we expect it. Realising the complexity and variability of the conditions, and the consequent precariousness in the measuring, we are content if the two sets of measurements roughly correspond: we are content with a lower degree of reliability. The reliability of a test is mathematically determined by giving it twice to the same children and finding the correlation between the two sets of scores. If the correlation is perfect the reliability is perfect; if the correlation is zero the reliability is zero. The coefficient of correlation so obtained is known as the coefficient of reliability. No test is regarded as quite satisfactory unless the coefficient of reliability reaches ·8 or ·9; and no test is considered of any scientific value if the coefficient of reliability falls below ·6.

There is another and a broader meaning to reliability: the word indicates the validity with which any particular marks or measures may be used as a basis of inference. A test may be quite reliable in the narrower sense and wholly unreliable in the broader sense. We may measure the heights of a group of boys of 11 years of age, measure them again the next day and get identical results. That group is our sample, and in the test of the sample we have perfect confidence. But have we confidence in the sample? What certainty have we that if we measure another group of boys of the same age we shall get the same results? Shall we get even the same average? More important still, if it were possible to test the whole population of boys of that age, would the essential values be the same as

those of the first measurement? What, in a word, is the reliability of the first measurement?

One thing is certain: the average of a group of children taken at random will differ from that of another group taken at random. Yet though the mean varies from sample to sample it does not vary lawlessly. What, then, is the law of variation? Can we from the mean of one group infer the means of other groups? If so, how many groups must we examine? how many means must we measure? Or is there anything in the data supplied by the testing of one sample which gives us a clue to the means of other samples, and in the last resort a clue to the mean of the whole group of persons or things of which the sample is a representative part? Fortunately, there is. A good sample, a sample chosen at random and large enough to display the properties of the whole group—properties which we may presume to be those of normal distribution—is able to tell us of things outside itself. It points to certain possibilities in other groups; it points to certain likely values in the one large inclusive group.

This internal evidence of reliability in the sample is to be found in its size and its shape. By its size I mean the number of units of which it is composed, and by its shape I mean the manner in which these units may be arranged around the average. We feel more confident in our findings when we measure a thousand single specimens than when we measure only ten; and we feel still more confident if the individual measures cluster closely round the mean. For if they do not cluster closely round the mean, but are widely and irregularly disposed, we know not what vagaries we may meet in the next sample

A good illustration of the evidence afforded by the shape of the sample is given by Mr. Elderton and his sister in their *Primer of Statistics*. They set out to find the average number of heads that would result from tossing six coins twenty times. In this case the true average is already known; for it is manifest that the most likely number of heads is three. The first series of 20 tossings yielded an average of 2·95, the second an average of 3·05, and the third an average of 3·05. The experimenters continued thus until they had found 45 of these means. Then these means were tabulated and treated as though they were simple measures. Their average was 3, and their standard deviation ·28. Could this degree of dispersion have been inferred from the first series of tossings without going to the trouble of repeating the series 45 times? As a matter of fact it could—not with certainty, it is true, but with a reasonable measure of probability. The standard deviation of the first series of tossings (1·16) divided by the square root of 20, the number of tossings, gave ·27, which agreed closely with the result obtained from actual experiment.

Instead of the standard deviation, however, the measure of distribution usually chosen is the "probable error." The term dates back to the days when deviations were called errors, and when "curve of error" was the customary name for the normal curve of distribution.

The probable error or probable deviation is not more likely to occur than other deviations, but is that deviation which is just as likely to be exceeded as not to be exceeded. Fig. 29 represents the bell-shaped curve of normal distribution, with the base

line indicating the scale of values and the height of the curve the frequencies with which those values occur. The area between the curve and the base represents the total number of cases. The mean, *m*, is in the middle. The probable error is that distance which when marked off from the mean in both directions determines half the number of cases. Vertical lines have been drawn with the distances

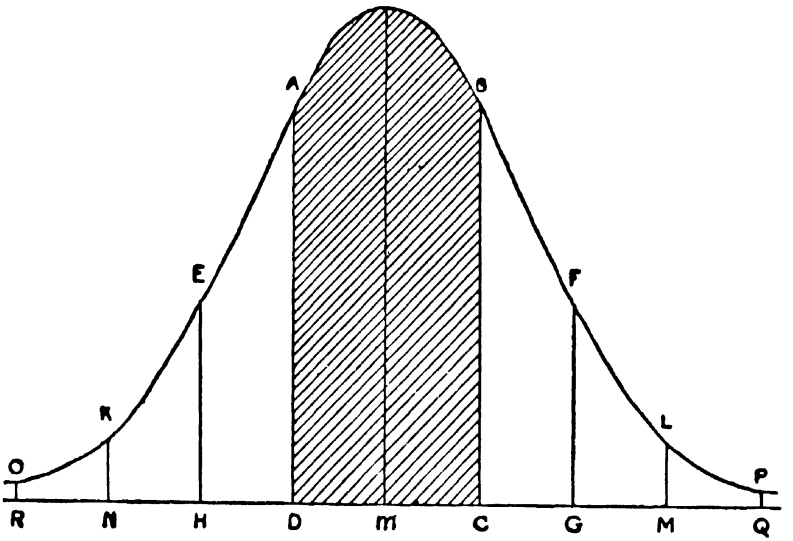


FIG. 29.

between them equal to the probable error. The shaded area A B C D, therefore, takes up half the entire surface: half the measures fall within it. Nine measures out of every eleven fall within the area E F G H (erected on ± 2 P.E.), 21 out of every 22 fall within the area K L M N (erected on ± 3 P.E.), and 993 out of every 1,000 fall within the area O P Q R (erected on ± 4 P.E.). Thus practically the whole of the measures are contained

within a range of four times the probable error on either side of the mean.

Probable error will be recognised as another name for the quartile deviation, or Q, or the semi-inter-quartile range. It has become customary, however, to use the term quartile deviation when speaking of the distribution of simple measures and to reserve the term "probable error" for the deviation of derivative measures such as the mean, the standard deviation, or the coefficient of correlation. For it is clear that these secondary measures can be arranged in the form of a distribution surface in precisely the same way as the primary measures from which they are derived. We can have a mean of means, and a standard deviation of standard deviations. The probable error of a calculated mean is the quartile deviation of all calculated means from the true mean; and the probable error of a calculated coefficient of correlation is the quartile deviation of all calculated coefficients from the true coefficient.

The simplest way to find the probable error of a series of measures is to count the measures; that indeed is the only way if we wish to be certain that $\pm$ P.E. will include exactly half the number of cases. But the modern custom is to derive the probable error from the standard deviation (σ). When the curve of distribution is normal the probable error is equal to the standard deviation multiplied by ·67449; and since the natural phenomena we are concerned in measuring may be assumed to conform to the normal curve with sufficient closeness to preclude the possibility of serious inexactitude, the probable error of simple values is in practice calculated from the formula P.E. = $\cdot 67449\sigma$, and the

probable errors of secondary values from the following formulæ:

$$\text{P.E. of a mean} = \cdot 67449 \frac{\sigma}{\sqrt{N}}$$

$$\text{P.E. of } \sigma = \cdot 67449 \frac{\sigma}{\sqrt{2N}}$$

$$\text{P.E. of } r = \cdot 67449 \frac{1 - r^2}{\sqrt{N}}$$

The probable error of a mean is sometimes called the probable error of sampling, because it is an index of the trust we may place in the sufficiency of the sample. In the Elderton experiment described above the standard deviation of the means was ·28, and their quartile deviation or probable error ·67449 × ·28, or ·19. This signifies that the odds were equal that any particular mean that was calculated would fall between 3 — ·19 and 3 + ·19; that is, between 2·81 and 3·19. When the range of values was doubled the chances were more than doubled; they were in fact 9 to 2. And when we state the chances of a calculated mean falling within a certain range of the true mean we at the same time state the chances of the true mean falling within a certain range of the calculated mean; for the distance from London to Bristol is the same as the distance from Bristol to London. Stated generally, the chances that the true mean lies within

± P.E. are 1 to 1
± 2 P.E. are 4·5 to 1
± 3 P.E. are 21 to 1
± 4 P.E. are 142 to 1
± 5 P.E. are 1310 to 1
± 6 P.E. are 19,200 to 1.

So important a function is the probable error of a correlation coefficient that I will illustrate and discuss its use in some detail by reference to an experiment which I carried out in the correlation of certain series of numbers obtained by chance. I took 40 dominoes, printed on the back of each a distinguishing letter of the alphabet, placed them all in a box and shook them up. Then I extracted them one by one. The first to come out was F; this I marked 1 on my list. The next was S, which I marked 2. Proceeding thus I secured a series of 40 numbers, indicating the random order of selection. By putting the dominoes back in the box and repeating the procedure I secured a second series of numbers. The F this time was no. 34, and the S no. 16; the other letters too were all different in rank. I proceeded thus until I had secured ten series of 40 numbers each. Now since the order of the items in each of these series was determined by chance, the true correlation between any two series was zero. To discover how closely the calculated correlations would correspond with what was known to be the true correlation, I set about finding by the method of ranks the correlation between every possible pair; and since there were 45 pairs I secured 45 coefficients. They clearly hovered about zero, 21 being on the positive side and 24 on the negative side. The exact arithmetical mean was ·007. The two extreme values were ·41 and —·35. The standard deviation of the correlations was ·18 and their probable error ·12. Verifying this by counting, I found that 21 out of the 45 fell between ·12 and —·12, and two were precisely ·12. The full distribution is displayed in the histogram given in Fig. 30.

If the formula for probable error is valid, it is not necessary to calculate 45 correlations to find the probable error: to calculate one is sufficient. The first correlation I worked out (that between the first and second series) amounted to —·26, and its probable error was ·10. There is, therefore, a

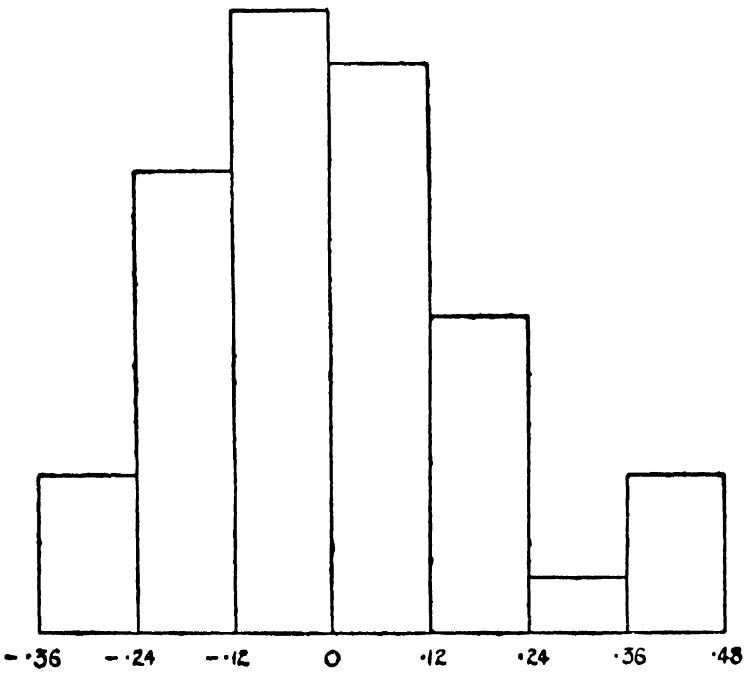


FIG. 30.—DISTRIBUTION OF THE CORRELATION COEFFICIENTS OF A CHANCE SERIES.

discrepancy between the two estimates of the probable error, the one secured by empirical methods being the larger by ·02. It does not follow, however, that it is the more accurate. It is in fact more than likely that if I had carried the experiment further, and calculated more correlations, I should have found the probable error getting nearer and nearer

·10. The empirical estimate was evidently unduly weighted by the unusual number of positive correlations that fell beyond three times the probable error.

We find precisely the same vicissitudes in the tossing of coins, or indeed in any of our dealings with chance. I may find that by tossing a penny 20 times I get 15 heads and 5 tails; but this does not alter the fact that the most probable incidence in 20 tossings is 10 heads and 10 tails. And a betting man would not hesitate to stake his money on that probability against any other. If the result falsified his reckoning he would not be surprised: he would merely be annoyed. Nor would his confidence in the line of reasoning be shaken: he would bet in exactly the same way again. So with correlations. If we compute the correlation between two sets of variates which are known to be independent, whatever correlation we may mathematically find is due to chance. But this does not mean that the calculated correlation will be nought: it means that a group of such calculated correlations will have nought as their central tendency. They will vary from that central point with a probable error or quartile deviation of ·15 if there are only 20 measures on which each coefficient is based, and of ·02 if there are as many as a thousand. With 40 measures, as in my experiment, the probable error is ·10. It is also approximately the probable error calculated by formula from my first correlation coefficient, —·26. The complete statement would appear thus: $r = -·26$, P.E. $= \pm ·10$; which means that the chances are even that the true correlation lies between —·36 and —·16; they are 9 to 2 that it lies between —·46 and —·06; they are 21 to 1 that it lies between

—·56 and ·04. But if it lies between —·56 and ·04 it may be zero, or it may even be positive. Thus the evidence afforded by the first coefficient, —·26, while presumptive evidence of a negative relationship between the two series, is not conclusive evidence of any connection whatever. Hence the common-sense rule of statisticians not to accept a correlation coefficient unless it is at least three times as large as the probable error. Indeed the more cautious insist on its being at least four times as large.

It will be observed that the highest chance correlation I obtained was ·41—a value which is more than four times ·09, its probable error. This single estimate would, therefore, be taken by the strictest of statisticians as strong proof of positive relationship between the two series; and yet, as we know, there is no such relationship. The fact is that chance may bring us anything, if we persist long enough it would even bring a perfect correlation between two sets of numbers chosen at random. If I have the patience (and the longevity) to go on shuffling my 40 dominoes and picking them up at random over and over again I shall ultimately—perhaps in a thousand years, perhaps to-morrow—draw a series which is the exact duplicate of my first series. The odds are tremendously against its happening at any particular time, or even within any particular time, but happen it will in the long run; just as every man is certain to be bitten by a mad dog, if he lives long enough. But as one bite one man comes on an average once every twenty million years (or thereabouts), we face the prospect with fortitude. The fact I wish to make clear is that any freak event of

this kind is just as likely to happen now as at any particular time that may be named beforehand. A perfect correlation between two independent series can happen—it is within the realm of possibility; and it is just as likely to happen the first time as at any other time that may be specified. My first correlation, instead of —·26, might have been 1, or — 1. The probable error would not have saved me; for in this case it would have been nought. It follows that even with the safeguard of a probable error tagged on to its tail a correlation coefficient must be treated with caution. We never can be quite sure. It is just as well to repeat the calculation a few times to see what happens.

The probable error is for all that a very valuable index of reliability—or of unreliability—and the reader who has any traffic with correlations should always append this index. Professor Karl Pearson and his students have saved us the bother of computing from formula by giving us a table of probable errors already computed. Others have constructed this ready-reckoner for themselves; and others are content to borrow. I am one of the borrowers. Table VI is quoted from Rugg.

In the example mentioned in the preceding chapter the correlation between two sets of mental tests given to 38 pupils was found to be ·74, calculated from the product-moment formula. This value of r must be considered along with its probable error, which from the formula $\cdot 67449 \frac{1 - r^2}{\sqrt{N}}$ (or from Table VI) is equal to ·049 or nearly ·05. Hence it is as likely as not that the true coefficient

TABLE VI

Probable Error of the Coefficient of Correlation

Number of Measures.	r						
	0·0	0·1	0·2	0·3	0·4	0·5	0·6
20	1508	1493	1448	1373	1267	1131	0965
30	1231	1219	1182	1121	1035	0924	0788
40	1067	1056	1024	0971	0896	0800	0683
50	0954	0944	0915	0868	0801	0715	0610
70	0806	0798	0774	0734	0677	0605	0516
100	0674	0668	0648	0614	0567	0506	0432
150	0551	0546	0529	0501	0463	0413	0352
200	0477	0472	0458	0434	0401	0358	0305
250	0426	0421	0409	0387	0358	0319	0272
300	0389	0386	0374	0354	0327	0292	0249
400	0337	0334	0324	0307	0283	0253	0216
500	0302	0299	0290	0274	0253	0226	0193
1000	0213	0211	0205	0194	0179	0160	0137

Number of Measures.	0·65	0·7	0·75	0·8	0·85	0·9	0·95
20	0871	0769	0660	0543	0419	0287	0147
30	0711	0628	0539	0444	0342	0234	0120
40	0616	0544	0467	0384	0296	0203	0104
50	0551	0486	0417	0343	0265	0181	0093
70	0466	0411	0353	0290	0224	0153	0079
100	0391	0345	0294	0242	0187	0128	0066
150	0318	0281	0241	0198	0153	0105	0054
200	0275	0243	0209	0172	0133	0091	0047
250	0246	0218	0187	0154	0118	0081	0042
300	0225	0199	0170	0140	0108	0074	0038
400	0195	0172	0148	0122	0094	0064	0033
500	0174	0154	0132	0109	0084	0057	0029
1000	0123	0109	0093	0077	0059	0041	0021

of correlation lies between ·69 and ·79. The chances are 9 to 2 that it lies between ·64 and ·84, 21 to 1 that it lies between ·59 and ·89, 142 to 1 that it lies between ·54 and ·94, and over a thousand to 1 that it lies between ·49 and ·99. This may therefore be taken as clear evidence of the existence of a fairly high degree of correlation between the two tests in question; that is, there is a distinct tendency for pupils who do well in the one test to do well in the other, and for pupils who are below the average in the one to be below the average in the other.

There is one principle of correlation which the teacher would do well to bear in mind. It is this: the more homogeneous the group tested, the lower will be the coefficient of correlation. When the pupils differ but little from one another in the ability measured, small errors of measurement are relatively more disturbing, and may even falsify the order of merit. It follows that we get clearer evidence of correlation if we take as our sample an age-group rather than a class.

Every natural group—every group which when measured in respect of a given quality conforms to the normal curve of distribution—has its own natural unit of measurement in respect of that quality. That unit is an index of the mode of dispersion. Some follow Galton in favouring the probable error as the unit; others of a more modern school favour the standard deviation. Each plan has its advantages. Where it is possible to assign absolute marks the standard deviation is perhaps the better; where it is possible to do no more than rank the performances in order of merit the balance of advantage lies with the probable error. Whichever of these

methods we adopt, the starting-point of our measurements must be not the best specimen, nor yet the worst specimen, but the average specimen, —either the median or the mean. Instead of beginning at the beginning we begin at the middle. For the middle is the stable point of the group—its centre of gravity—the capital of a country with shifting boundaries. The normal curve of distribution has no limit at either end: it extends indefinitely in both directions, getting nearer and nearer to the base-line but never touching it. At each extreme a door is left open for possibilities. We can never tell when we have gauged the greatest height of intelligence or sounded the lowest depth of stupidity.

We will suppose that we have taken the probable error as our unit and have graduated in Fig. 31 the base-line of a curve of frequencies in terms of that unit. The positive half only appears in the diagram, for if the distribution is normal, as it is assumed to be, the negative half will be an exact duplicate. The diagram reveals the rate at which the numbers diminish as they recede from the average. The 10 per cent. of cases immediately above the mean correspond to ·376 P.E., the next 10 per cent. to ·402 P.E., the next to ·470 P.E., the next to ·652 P.E., and the highest group containing 9·9 per cent. of the cases correspond to 2·700 P.E. This last group is incomplete, as a minute part of the area lies beyond this boundary. Since the successive multiples of the probable error represent steps of equal difficulty, we cannot judge the merits of performances as inversely proportional to the number of candidates who succeed in them. It is

much easier to rise from the 50 percentile, or median, to the 60 percentile, or the tenth per cent. above the median, than from the 80 percentile to the 90 percentile. The actual ratio facility is 376 to 652.

Use was made of these facts by Mr. Cyril Burt in constructing his scale for handwriting. He collected a number of specimens for each age-group and arranged them in order of merit. "From the whole

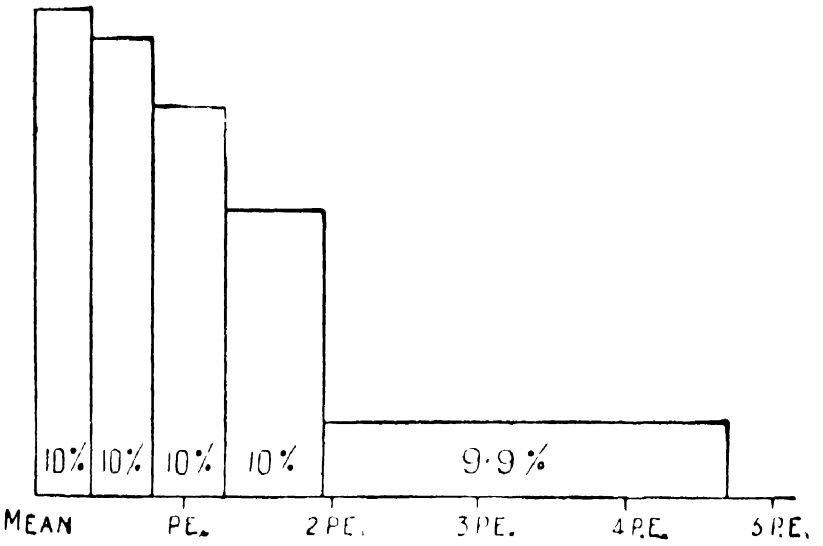


FIG 31

series thus graded nine specimens were then extracted: (1) the first or best; (2) the last or worst; (3) the middle specimen or 50th per cent. in the whole series (the 'median'); (4) the middle specimen of the upper half of the series (the 25th per cent. or 'upper quartile,' marking a distance of 1 × p.e. above the median or average); (5) the middle specimen of the lower half of the series (the 75th per cent. or

'lower quartile,' marking a distance of 1 × p.e. below the median); (6) the 10th per cent.; (7) the 91st per cent.; (8) the 3rd per cent.; and (9) the 98th per cent., representing respectively 2 × p.e. above and below, and 3 × p.e. above and below, the median.

"We thus secure scales for measuring individuals of a given year in terms of the average and variability of their own age-group. Since the unit of variability is the same throughout the scale, it may be assumed that the intervals between successive specimens are equal."[1]

The reader will now realise that a unit of deviation is comparable with mental age. Each is a unit of measurement derived empirically from a homogeneous group—generally an age-group; but while one is based on the distance between similar points in two successive groups, the other is based on the distance between two points within the same group. It is therefore possible to use the unit of variation where the unit of mental age is quite inapplicable. Since it has never been possible to carry the scale of mental ages beyond the age of 15 or 16, the mental-age system has no standards at all (except certain arbitrary standards arrived at by guesswork) by which may be estimated the intelligence of people who are 16 or over and who are above the average of the sixteen-year age-group.

If, however, we are right in assuming that all men over 16 form in respect of intelligence a natural group, just as all men over 20 form in respect of stature a natural group, then we have in the standard deviation, or the probable error, of that group a

[1] *Mental and Scholastic Tests*, p. 309.

unit of measurement by which any individual in the group may be scientifically assessed.

The reader is now in a position to put meaning into the bald statements which I now give of certain correlations which have been calculated from the results of some of the group tests described in this book. The coefficient of reliability of the Columbian Tests is ·93, with a P.E. of ·01; and that of the Chelsea Tests ·90, with a P.E. of ·01. The correlation between the Columbian and Chelsea Tests is ·86, with a P.E. of ·02.

Those who wish to make a more complete study of the theory of correlation will find excellent material in Nunn's *Exercises in Algebra*, Vol. II, Sec. ix (Longmans). Perhaps the best book of all for the teacher is Rugg's *Statistical Methods Applied to Education* (Harrap), where the subject is treated with much lucidity and in considerable detail. *The Essentials of Mental Measurement*, by Brown and Thomson, is a standard book of its kind, and to the advanced student it can strongly be commended; but it is far too difficult for the beginner. The very beginner, who has no knowledge whatever of statistical methods, cannot do better than start with reading an excellent little book called *A Primer of Statistics* by W. P. and E. M. Elderton (Black).

CHAPTER XX

THE USES OF GROUP TESTS

THE essential difference between an examination and a mental test is that one points to the past and the other to the future. An examination is intended to reveal what the pupil has done in the years that are gone; a mental test is designed to foretell what he may do in the years to come. An examination is an inventory, a mental test a prophecy. It thus happens that where mental tests have displaced an examination it has always been an admission examination rather than a leaving examination. Mental tests have for many years now guarded the door of the school for mentally defectives all over the civilised world; they have begun to guard the door that admits children from the elementary school to the secondary school; they have even begun to guard the entrance to the university.

The pioneer of mental testing at the university is Professor Thorndike, of the Teachers' College, Columbia University, New York. The candidate for admission to the university may, if he likes, take a series of mental tests in lieu of the ordinary entrance examination. And he often does like. The others—those who elect to take the ordinary examination—are also given the mental tests; not for purposes

of selection, but for purposes of comparison. It is thus possible to say which mode of selection winnows the candidates the more successfully—which method gives the better index of ability to profit by the courses of study provided by the university. Let the Dean of the College speak for himself[1]: "Fortunately it is possible to determine with scientific accuracy whether or not the mental test is a useful addition to our academic machinery. If it turns out, during a series of years, that the correlation between the marks received on the mental tests and the collegiate work of the students is distinctly higher than the correlation between the results of other types of entrance examinations and the college work, it would seem to be clear that the new plan of admission affords the best index that we have of the ability of a boy to carry college work. The correlation between the work of the entire freshman year for the students who entered by the new plan and their marks on the mental test is + 0·65. The most reliable data available indicate that the highest correlation that can be expected between the work of the freshman year and the results of the usual college entrance examinations is about + 0·45. This latter figure has been obtained not only from a statistical study of our own freshmen but from similar studies in another institution."

Besides the two grounds of selection referred to above there is another ground which has been adopted by certain American universities—the reports on the candidates' work at the high school. This source of information is not ignored at the

[1] Quoted from a Report by Dr. D. E. Hawkes, Dean of Columbia College, in the *Journal of Educational Research*, for Sept. 1921.

Columbia University; indeed no candidate is allowed to take the mental tests unless his school record is satisfactory. And by the method of correlation mentioned in the Dean's report it has been proved that although the school record has greater prognostic value than the entrance examination, it has less prognostic value than mental tests. Mental tests head the list each time.

Although Professor Terman gives a less favourable account of the present efficacy of mental tests at the Western universities,[1] it is to mental tests that he looks for deliverance from the patent evils of the ordinary mode of selection. By the ordinary mode at least 10 per cent. of failures are let through—students who have not the brains to graduate—students who not only exclude abler people, but lower the general level of scholarship, set a bad example to others, waste the time of the professors, and clog the educational machinery.

Terman found the correlation between class marks at the university and class marks at the high school to be 0·54, while the average correlation between intelligence marks and university examination marks was only 0·49. Thus the verdict of the Stanford University is at variance with the verdict of the Columbia University. Terman accepts the view that correlations between intelligence scores and success in study are considerably lower at the University than at schools for younger pupils. In other words, the native quality of a pupil's mind is a safer earnest of scholastic success when he is young than when he is old. The main reason that Terman gives for this is that character counts for

[1] *School and Society*, April 23, 1921.

more at college than at school; at college a student is thrown more on his own resources, and has to withstand by his own self-restraint the distracting influences of sport and social life. And mental tests leave character out of account.

Curiously enough, however, Terman forgets to mention the obvious fact that the differences in the store of knowledge acquired by the individuals of a given group are accentuated as those individuals grow older. And the success of a man at the university depends no less on his store of knowledge—on the grounding he has received—than on his natural gifts. Even the authorities of the Columbia University would refuse to admit a raw and uneducated cowboy, however great his natural intelligence. They insist on a guarantee of book-learning as well as of mother wit.

The discrepancy, however, between the Stanford verdict and the Columbia verdict still remains unexplained. Columbia on the mathematical evidence places mental tests above the high-school records; Stanford on the same kind of evidence places the high-school records above mental tests. The explanation of course lies in the fact that there are mental tests and mental tests. The more one studies the various group tests in use in America, the more one realises that they do not test one constant factor nor yet one constant group of factors. The favourite criterion of validity is success in some particular sphere of life. The American army tests were regarded as valid tests of intelligence to the extent that they measured future success in soldiering; the group tests for schools are regarded as valid to the extent that they measure

a pupil's capacity to profit by schooling; the group tests for universities are regarded as valid to the extent that they measure the likelihood of a student's doing reasonably well at his university examinations. This view is expressly put by Haggerty in the Manual of Directions for his Intelligence Examination; for he says: "However we may technically define intelligence, what we mean by it here is that capacity which enables a pupil to profit by the work of a good school, to recognise the meaning of words and sentences, to remember important facts observed and learned, to discriminate essential from unessential materials, and to infer correct conclusions from presented data."

It seems highly probable, therefore, that Terman and Thorndike secured different results at their several universities for the simple reason that they used different tests. What Terman's tests were he does not state, but Thorndike's are published for all the world to see. They are incomparably more difficult and more searching than the army tests; and although the majority of them fall under one or other of the general types of test already described, others, as will be seen from the following quotation, presuppose an academic training. The candidate is asked to say whether these statements are true or false:

"$\text{Sin } x = \pm \frac{1}{\sqrt{\cot^2 x + 1}}$

"The velocity of a wave motion equals the product of the wave length by the frequency.

"The annelids are higher animals than the molluscs.

"Arterial blood has about 20 c.c. of oxygen dissolved in every 100 c.c. of blood."

Aristotle with all his wisdom could have answered none of these questions. We are forced to conclude that the group tests of intelligence at present in use in America are not tests of an intelligence that is equally at home in every department of life: they have a slight vocational bias. Even when the form of the tests is the same, the contents vary with the line of activity which the candidate intends to follow.

The use of group tests of intelligence for selective purposes has not been slow in reaching this side of the Atlantic. Education authorities in England have begun to use them for the award to scholarships to secondary schools. They are being extensively used for the same purpose in Germany, especially at Berlin, Hanover and Leipzig. At Hamburg 1,000 children are tested each spring, and the best of them transplanted to higher nurseries of learning.

The system of selection by mental tests is not confined to schools and colleges: it has invaded municipal and business life. The Civil Service has partly adopted it, and a few large business firms in London and elsewhere seem to have wholly adopted it.

But mental tests are not only selective; they are diagnostic as well. Dr. Hawkes points out that the tests used at the Columbia University have greatly helped him to arrive at a diagnosis of academic maladies. "A student who has a poor academic record and a low mental-test grade generally needs very different treatment from the student whose record is poor but whose mental-test mark is high.

The same principle applies to the "retarded" class which is becoming so common a feature in the organisation of the elementary school. In bygone days it was called the dunces' class, and the time was spent in grinding at the three R's. Nowadays it is given an inoffensive, or even an honourable, name, and the bulk of the time is spent in practical and concrete work. It should be noted, however, that a class of this kind generally contains two distinct types of pupils. Some of them are there through lack of opportunities and some through lack of brains. Intellectually they are all ill-nourished, but some are ill-nourished because they have been starved, others because they have weak digestions. The former can be given more food of the same kind as before, the latter need food of a different kind. To the former the class should be a clearing house, to the latter a permanent home. And not only in the "remove" class is it necessary to distinguish between the dull and the retarded, but in every class. It is true that the retarded are generally dull as well: but it is not always so. And the surest way to find out when it is not so is by the giving of mental tests.

CHAPTER XXI

MENTAL TESTS AND SCHOOL ORGANISATION

THE orthodox basis of school organisation is the attainments of the scholars. Those boys and girls who have arrived at the same stage in the course of study provided by the school are put in the same class and taught by the same teacher. And this seems so eminently reasonable a plan that to challenge its adequacy savours of captiousness. Yet challenge it we must; for it is patent to the discerning mind that in the elementary school at least the system has not worked well in the past, and is not working well at the present.

Until the real way in which the abilities of children are distributed was revealed by the application of mental tests, the prevailing belief was that there were two broad classes of children, the normal and the mentally defective, the latter forming a small group sharply divided off from the former, and, especially since the foundation of special schools, constituting a negligible factor in the school economy. The vast majority of the children were normal, there was nothing wrong with them, and they all therefore ought to pass through the standards at a uniform rate of one standard per annum. If a healthy child failed to keep step with the procession, the failure was ascribed to bad teaching. If his natural pace

would carry him in advance of the procession, he was suspected of an undesirable precocity, and made to learn the salutary lesson of keeping pace with his comrades. And there are teachers still living who, remembering dimly the days when the passage from standard to standard was an annual event determined by an annual examination, cling to the belief that the bulk of the children really did make that pilgrimage through the school, arriving at last at the top standard with a store of knowledge and accomplishments which would have astonished the modern pedagogue. What these children were like when they did arrive at the haven of this seventh standard I have no means of showing; but I can show that the number of those who actually did arrive was pitiably small.

Fig. 32 gives in diagrammatic form the distribution of children among the various standards in London board schools in March 1886. The columns are proportional to the numbers given in the published records. At that date annual government examinations were in full vogue and vigour; and here we see a typical result of the classification of scholars based on their success at a stringent educational test applied once a year. The decline in numbers from standard to standard was extraordinarily steep; for all the forces made for retardation and none for acceleration. A child who was absent from the annual examination because he was ill, or because he fell short by a few weeks' training of the required level, had to remain for another whole year in the same class; while the child who did the whole work in the first three months had to mark time for the rest of the year.

If we leave out of the reckoning the infants below

Standard I, we find that half the children in the London elementary schools of 1886 were in standards I and II; and about half the children left before they reached Standard V. It is true that in those days they were not compelled to attend school after 13 years of age, but 30 or 40 per cent. of them did as a matter of fact remain beyond that age.

In the same diagram is shown the distribution of children in London schools in 1910, the last date

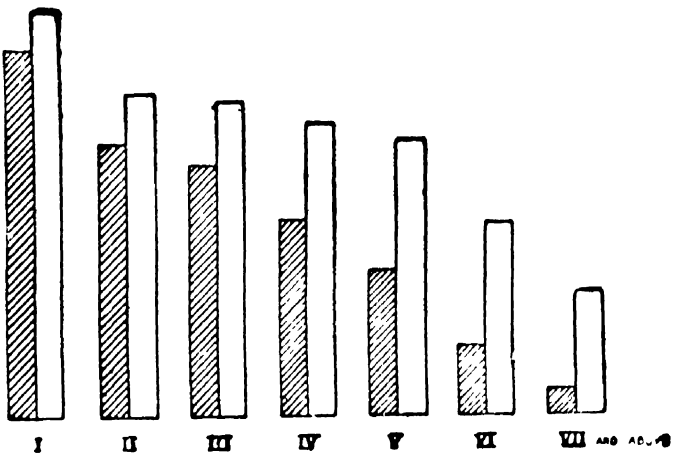


FIG. 32.—1886 (SHADED), 1910 (BLANK).

when statistics of this kind were published. It will be seen that by this time things had improved; but again there was an inordinate crowding of children in the lower part of the school and a glaring discrepancy between the theoretical number of children per standard and the actual number of children per standard. And this was nearly two decades after annual examinations by the Board of Education had been abolished. The truth is the tradition remained. Promotion was still yearly, and it was

still based on a scholastic examination. The only difference was that the head master set the examination instead of the government inspector.

The next step in London was the adoption of a scheme of double promotion. Children were promoted twice a year instead of once; and it was no longer permissible to have monster classes of Standard I and Standard II children and diminutive classes of Standard VI and Standard VII children. All the rooms had to be reasonably filled with children, and none had to be crowded. The result was a levelling of the size of the classes, and a blurring of the significance of standard. The top class in the school no longer consisted of the few stalwarts who had run the gauntlet of six or seven standard examinations. It was often as large as any other class in the school, and it not infrequently contained children who were unfitted to cope with the syllabus of work. And instead of 50 per cent. of the children reaching no higher than Standard IV, 50 per cent. now reach the highest class in the school, which sometimes is Standard VII and sometimes Standard Ex-VII. When the children are classified, however, by the stringent standards of former days, there is seen the same decline in numbers from standard to standard, as is abundantly witnessed by the figures in Mr. Burt's book on the *Distribution and Relations of Educational Abilities*. The decline, however, is much less rapid than of yore. The weaknesses of the system are less, but they are still there. They are still there because the basis of promotion remains unchanged: with rare exceptions promotion is to the present day solely determined by a scholastic examination periodically applied.

We may learn much from what has recently been happening in America. In the year 1909 Leonard P. Ayres published a notable book called *Laggards in our Schools*, a book which showed that about one-third of the pupils in the public schools of New York and other large cities were too old for their grades. They were too old even when a generous margin was allowed of one year beyond the recognised normal age for the grade. Administrators got alarmed and talked about the failure of the schools and the waste of public money. The alarm spread to other parts of America, and investigations were made on an extensive scale. Professor Strayer secured data on the amount of retardation in 318 cities, and he found that while one child out of every three was retarded, only one out of every twenty-five was accelerated. The inference drawn from these researches was that the older children were being unjustifiably kept back—that they were in fact normal children who did not progress at the normal rate because they did not receive normally good teaching. The result was a general "hustle." Attempts were made to get the ages right by forcing up the older children—the "laggards" as they were called. But it was soon discovered that when the ages became right the grades became wrong. Grade x could no longer be called grade x, if judged by the accepted standard of attainments for that grade. The final outcome seems to have been a reversion to the older classification; so that at the present day things are much the same as they were before the publication of Ayres's celebrated book. At any rate Terman says so. He contends that the evils of retardation have not been remedied, for "the

number of school laggards has decreased but little, and their needs are almost as little provided for as before the campaign on their behalf began."[1]

The fact is these earlier investigators had discovered a disease, but had diagnosed it wrongly. They assumed a closer correspondence between real age and scholastic age than in point of fact existed. They assumed that it was the older children who were the laggards, whereas the real laggards were the younger children.

The belief in the level-age class is a natural outcome of the traditional scheme of school organisation. If a normal child should normally reach a certain standard at a given age, it is quite reasonable to regard a wide range of ages in that standard as a mark of either bad organisation or bad teaching. This view, however, is flatly at variance with statistical laws. A wide range of ages, far from being symptomatic of bad organisation, is the invariable sign and seal of good organisation. For a careful measurement of scholastic abilities shows that these abilities tend towards that peculiar form of crowding and scattering which is known as normal distribution. This characteristic is brought out quite clearly in Fig. 33, which represents the distribution of Grade VI children in the public elementary schools of Connecticut in 1903.[2]

The range of ages in the diagram is unusually large because the number of children is unusually large. With a smaller number there is a smaller range; but even in an ordinary class of 40 the

[1] *The Intelligence of School Children*, p. 111.

[2] The numbers are taken from Strayer and Thorndike's *Educational Administration*, p. 4.

natural range of ages may extend over four or five years. So regular a curve is exceptional. The distribution of ages in English schools has never been so symmetrical as this; yet never has the system of classification and promotion been so bad as seriously to disturb the natural order of things. The trend is everywhere towards normality of distribution, as will be seen from a study of the numerous dia-

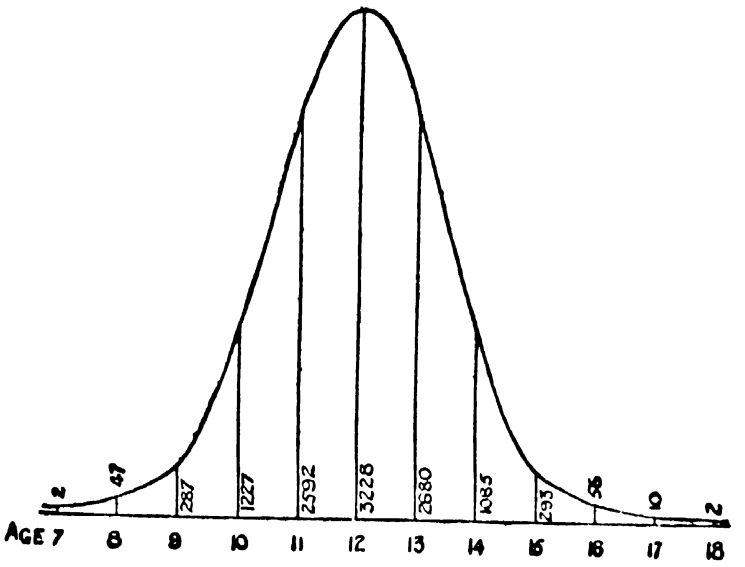


FIG 33.

grams in Mr. Burt's book. It follows that if in a class the scholastic ages are the same the birthday ages will be mixed, and if the birthday ages are the same the scholastic ages will be mixed; and to grumble about a wide range of ages in a class is as unreasonable as to complain that the stature of half the population of England is below the average. If the central tendency of the class is right, let the deviates be what they may, so long as they are symmetrical.

But the trouble is that the deviates are not symmetrical. They are not symmetrical because the prevailing belief in a uniform calendar age constantly militates against the securing of a uniform scholastic age; for it gives rise to the universal practice of promoting the older and stupider children faster than their ability warrants, and promoting the younger and brighter children more slowly than their ability warrants. The consequence is that in the ordinary class the younger children gravitate towards the top of the examination list and the older children towards the bottom; which is another way of saying that the real laggards in the school are not the older children, but the younger children.

The defects of our present system of organisation do not all spring from a belief in uniformity of age in the ideal class—a belief that is almost unconscious, but endued with all the stubbornness that belongs to the unconscious; some of them spring from the rigidity of the practice of promoting at fixed intervals; and some of them spring from the ignoring of mental age.

Classification on a basis of birthday age is fairly common in our infant schools. It is not a good basis, but in the absence of evidence of intelligence and attainments it is the only possible basis. But then a classification by the birth certificate is less harmful in the infant school than anywhere else; for the absolute difference between child and child is as yet small. It is the beginning of a race, before the runners have had time to scatter.

Classification by mental age has yet to be tried. Those few schools that have ventured cautiously along this new road bring us good tidings. In a dreary

quarter of a great city there stands one of the biggest elementary schools in England. To the girls' department a new head mistress was appointed in 1913. For the previous decade not a single scholarship had been gained at the school. The new head teacher had faith in mental tests, and classified her school according to the intelligence of the pupils. She had three currents of promotion, the speed of the current being determined by the intelligence quotient of the girls. The result was that during the next nine years as many as 33 scholarships were gained—scholarships that testify to at least two qualities on the part of the winners: natural intelligence and school attainments. Under the old classification the intelligence was there, but not the opportunity; the new classification brought the opportunity as well.

Terman has used the mental test as a weapon of attack on the organisation of American schools. After he and his students had tested a large number of school children he came to the conclusion that only about 50 per cent. of them were placed in the grade to which their mental age entitled them. About half of them were either accelerates or retardates—generally retardates. And the few accelerates were mainly the older and duller pupils, while the many retardates were the younger and brighter pupils.

There is little doubt that our school system favours the detecting of the dullard and the overlooking of the genius. It is easy for the brilliant boy to conceal his intelligence; it is difficult for the dunce to conceal his dullness. There is little doubt too that much nonsense has been talked about the danger

of pressing forward the brilliant boy. He is supposed to be a fragile creature with a delicate nervous system, and a brain peculiarly susceptible to injury from books; and his appetite for learning is supposed to be greater than his digestion, so that he has to be carefully dieted; his studies must not be too stimulating: they must be soft and bland and meagre. This boy is a myth. It is readily admitted that delicate children, children with unstable nervous systems, are to be found among all classes; but there is no reason to think that they are more plentiful among the brilliant than among the dull. All the statistical evidence goes to show that on the whole the supernormal child is healthier and stronger than the normal child, and the normal child healthier and stronger than the subnormal child. And as for over-pressure, it is not to be denied that that may exist among older boys lured by ambition, or goaded by a fear of disgrace; but younger boys may safely be trusted to resist the sort of brain-work that brings them discomfort. And where there is a victim of over-pressure he is far more likely to be a stupid boy than a brilliant boy.

Medical men, who used to be the alarmists in the matter of over-pressure, have recently been silent. Some of them, who have come under the influence of the New Psychology, assert that thwarted ambition at school may give rise to a "complex"; and a complex is a thing to be avoided at all costs.

If the question at issue were simply this: Are we to change the basis of school organisation and classify on mental age and nothing else whatever? we should have no hesitation in answering No! No

single basis, whatever it is, will serve our purpose. What a child has to do in school is to pursue school studies, and to ignore his stage of progress is to ignore the most obvious, and apparently the most pertinent, datum. Nor can we infer his scholastic age from his mental age, although the two tend roughly to go together. It is a curious fact that children of the same mental age but of different calendar ages are far from displaying the same educational aptitudes. It is found, as a general rule, that if a child of 8 with a mental age of 10 is taught in the same class as a child of 12 with a mental age of 10, then, in spite of a much shorter period of exposure to educational influences, the younger child is the better scholar. In fact it may be stated generally that the child with a high mental ratio is well suited for book-learning, whatever his calendar age may be; while the child with a low mental ratio never takes kindly to books. Fortunately he has a thick skin, and a fine system of defensive mechanisms.

If we aim at an ideal classification we cannot afford to ignore any source of information about our pupils. To limit the source to one, whatsoever that one may be, is to court failure—or at least to fall short of complete success. As a matter of fact the teacher often adopts more than one basis of classification: he considers other things besides attainments. He puts forward the old dull boy in the interests of the boy's self-respect; he keeps back the young bright boy in the interests of the boy's social training. So far he is right. He has seen the problem and attempted a solution; but it is not, I submit, the best solution. And the best solution

can never be reached until full weight is attached to the factor of mental age—the factor that has in the past been most generally ignored; and ignored for the reason that we have not known how to assess it.

But although no one basis of classification will suffice, it seems clear that of the two important rivals—mental age and scholastic age—precedence belongs to the more constant and the more potent. And mental age (or rather mental ratio, which, theoretically, never changes) is manifestly a more intrinsic and stable part of the pupil's mind than the knowledge he may gain or the habits he may form. Whatever we may do or say, his intelligence will prove a more potent factor than his store of knowledge in determining the broad lines of his career at school; and not in school only, but in life also. Character may of course come in and deflect these lines at any point, but so may illness, and many other things that disturb our calculations.

We arrive at the conclusion that the intelligence of the pupils is the primary, but not the only, basis on which a school should be organised. Other factors gradually insinuate their claims; but in the early stages at least intelligence should be the dominant factor. There should be distinct streams of promotion through the school—three is a convenient number—and the factor that should decide in which stream the pupil should be placed can be no other than his mental ratio or intelligence quotient. No other factor tells us so much about him; no other factor is so vital and so intimate, so fruitful a source of suggestion, so valid a ground of inference. It should prevail over every other consideration in settling the pupil's general status in the school com-

munity. It alone indicates the lines of cleavage between the intellectual levels of the pupils; and without the knowledge it supplies we are limited to a single line of progress through the school. And with a single line of progress through the school it is impossible to solve the problem of the old dull boy and the young bright boy.

Let it not be thought that in advocating a triple current of promotion I propose that each class or standard should appear in triplicate. The ordinary standards fit the middle stream; they do not fit either the fast or the slow. Besides, none but the very largest schools could supply enough children to fill the three channels. What I suggest for the school of ordinary size is that each class should have its three sections, the fast, the medium, and the slow; and each section have its own syllabus. The normal line of progress would be from a given section of one class to the corresponding section of the next. The method of solution which is rapidly coming into vogue in our schools is to collect the slow-going pupils into one class and give it a special syllabus. It is more of a pond than a stream. Two classes would be better; yet one is preferable to leaving the retardates mixed indiscriminately with the average and the accelerates. But the schools that have these special classes for the dull have no corresponding special classes for the bright; not because the need of these classes is not felt, but because they give rise to the suspicion of illicit coaching for scholarships. But there is no reason why it should. If a teacher wishes to break regulations he can do it just as well with a few candidates in his class as with many candidates in his class; and perhaps better.

That the evils of the orthodox system of school organisation are crying aloud for redress is obvious from the drastic remedies that our schools are adopting. The aim of every system of school organisation is to secure homogeneous teaching units called classes, or grades, or standards. But our teachers are fast realising that the classes are never homogeneous, and that whatever the alleged teaching unit may be the learning unit can never be anything else but the individual pupil. Hence the cry for individual work everywhere. Hence the vertical classification in our infant schools, where there is no promotion from class to class, but only promotion from exercise to exercise; and the Dalton plan in our senior schools where the pace is set not by the class, but by each separate pupil for himself. Yet even in these extreme reactions against collectivism, the class as a teaching unit, as distinct from the class as an administrative unit, is far from being abolished. Collective lessons are still given, either to the ordinary class, or to a section of the class, or to special recombinations of classes that have temporarily been dissolved.

The movement towards individual work is all making in the right direction; it has cut adrift from the old rigid organisation and is dispelling the fiction that the children make a processional march through the school curriculum. I heartily believe in individual work, if taken with a due admixture of collective work. To adopt purely individual work would be to throw away the baby with the bath-water. But then nobody does adopt purely individual work. Collective work of some kind is inevitable; and collective work means classification

(good or bad); and good classification is based mainly on intelligence. Even in individual teaching itself a knowledge of the mental powers and possibilities of each separate pupil is the very secret of success. And nothing can give this knowledge so well as mental tests.

CHAPTER XXII

SPELLING DEMONS

OUR forefathers learnt spelling from a spelling-book. At any rate they tried to learn it. The book was built on some sort of principle, but on no sort of practice. It generally began with *a-b*, *ab*, and ended with polysyllabic words crawling like huge caterpillars across the page, words like intercommunication, antiparliamentarianism, and incommensurability. And a word like apothegm, which nobody used, was as dear to the heart of the maker of spelling-books as a word like wonderful, which everybody used; indeed dearer, for it was harder to spell.

But spelling-books, on both sides of the Atlantic, are things of the past; and spelling-bees no more improve the idle mind. No longer do we spell for display, but simply to avoid disgrace. We learn to spell not as many words as we can, but as few words as we can. The question for the classroom is: What words must we insist on our pupils knowing how to spell? The usual English answer is: Each child should be able to spell the words in his own vocabulary. But a child's vocabulary is not a thing made, but a thing in the making—a vague and indeterminate thing that shifts and changes and is never quite what it might be or what it ought to be.

The American answer is: A child should first learn to spell the commonest words in the language; and the commoner the word the more important is it that he should be able to spell it. Investigators have accordingly set themselves the task of finding the frequencies with which various words are used in speech, in letter writing, in school composition exercises, and in general literature; and they have compiled lists of spellings based upon those frequencies. Dr. Daniel Starch, for instance, has compiled a list of 2,626 words which he regards as constituting the ordinary working vocabulary of the educated man. He has a right to his opinion, for he has counted the words. Exactly how he has counted them is told in his book on *Educational Psychology*.[1]

The first list of this kind, however, and probably the best, was that of Dr. Leonard P. Ayres, whose research I have described in *Mental Tests*.[2] Through the kindness of Dr. Ayres I am here able to reproduce this complete list of the 1,000 commonest words in the English language. They are arranged in 26 groups named after the letters of the alphabet. All the words in a group are believed, on experimental grounds, to be of equal difficulty. In Table VII, I give the percentage of words in each group that should, according to the author, be spelled correctly by children in the various standards. They are given by Dr. Ayres in grades; but I have represented each grade by the standard that roughly corresponds with it. Grade VI, for instance, I have called Standard V; Grade V, Standard IV, and so on.

[1] *Educational Psychology*, pp. 328–33.

[2] *Mental Tests*, pp. 156–7.

AYRES'S SPELLING SCALE

A

Me do.

B

And go at on.

C

A it is she can see run.

D

The in so no now man ten bed top.

E

He you will we an my up last not us am good little ago old bad red.

F

Of be but this all your out time may into him to-day look did like six boy book

G

By have are had over must make school street say come hand ring live kill late let big mother three land cold hot hat child ice play sea.

H

Day eat sit lot box belong door yes low soft stand yard bring tell five ball law ask just way get home much call long love then house year to I as send one has some if how her them other baby well about men for ran was that his led lay.

I

Nine face miss ride tree sick got north white spent foot blow block spring river plant cut song winter stone free lake page nice end fall feet went back away paper put each soon came Sunday show Monday yet find give new letter take Mr. after thing what than its very or thank dear west sold told best form far gave alike add.

J

Seven forget happy noon think sister cast card south deep inside blue post town stay grand outside dark band game boat rest east son help hard race cover fire age gold read fine cannot May line left ship train saw pay large near down why bill want girl part still place report never found side kind life here car word every under most made said work our more when from wind print air fill along lost name room hope same glad with mine.

K

Became brother rain keep start mail eye glass party upon two they would any could should city only where week first sent mile seem even without afternoon Friday hour wife state July head story open short lady reach better water round cost price become class horse care try move delay pound behind around burn camp bear clear clean spell poor finish hurt maybe across to-night tenth

sir these club seen felt full fail set stamp light coming cent night pass shut easy.

L

Catch black warm unless clothing began able gone suit track watch dash fell fight buy stop walk grant soap news small war summer above express turn lesson half father anything table high talk June right date road March next indeed four herself power wish because world country meet another trip list people ever held church once own before know were dead leave early close flower nothing ground lead such many morning however mind shall alone order third push point within done body.

M

Trust extra dress beside teach happen begun collect file provide sight stood fix born goes hold drill army pretty stole income bought paid enter railroad unable ticket account driven real recover mountain steamer speak past might begin contract deal almost brought less event off true took again inform both heart month children build understand follow charge says member case while also return those office great Miss who died change wire few please picture money ready omit anyway.

N

Except aunt capture wrote else bridge offer suffer built centre front rule carry chain death

learn wonder tire pair check prove heard inspect itself always something write expect need thus woman young fair dollar evening plan broke feel sure least sorry press God teacher November subject April history cause study himself matter use thought person nor January mean vote court copy act been yesterday among question doctor hear size December dozen there tax number October reason fifth.

O

Eight afraid uncle rather comfort elect aboard jail shed retire refuse district restrain royal objection pleasure navy fourth population proper judge weather worth contain figure sudden forty instead throw personal everything rate chief perfect second slide farther duty intend company quite none knew remain direct appear liberty enough fact board September station attend between public friend during through police until madam truly whole address request raise August Tuesday struck getting don't Thursday.

P

Spend enjoy awful usual complaint auto vacation beautiful flight travel rapid repair trouble entrance importance carried loss fortune empire mayor wait beg degree prison engine visit guest department obtain family favour Mrs. husband amount human view election clerk though o'clock support does regard

escape since which length destroy newspaper daughter answer reply oblige sail cities known several desire nearly.

Q

Sometimes declare engage final terrible surprise period addition employ property select connection firm region convict private command debate crowd factory publish represent term section relative progress entire president measure famous serve estate remember either effort important due include running allow position field ledge claim primary result Saturday appoint information whom arrest themselves special women present action justice gentleman enclose await suppose wonderful direction forward although prompt attempt whose statement perhaps their imprison written arrange.

R

Forenoon lose combination avenue neighbour weigh wear entertain salary visitor publication machine toward success drown adopt secure honour promise wreck prepare vessel busy prefer illustrate different object provision according already attention education director purpose common diamond together convention increase manner feature article service injure effect distribute general to-morrow consider against complete search treasure popular Christmas interest.

S

Often stopped motion theatre improvement century total mention arrive supply assist difference examination particular affair course neither local marriage further serious doubt condition government opinion believe system possible piece certain witness investigate therefore too pleasant.

T

Guess circular argument volume organise summon official victim estimate accident invitation accept impossible concern associate automobile various decide entitle political national recent business refer minute ought absence conference Wednesday really celebration folks.

U

Meant earliest whether distinguish consideration colonies assure relief occupy probably foreign expense responsible beginning application difficulty scene finally develop circumstance issue material suggest mere senate receive *respectfully* agreement unfortunate *majority elaborate citizen* necessary divide.

V

Principal testimony discussion arrangement reference evidence experience session secretary association career height.

W

Organisation emergency appreciate sincerely athletic extreme practical proceed cordially character separate February.

X

Immediate convenient receipt preliminary disappoint especially annual committee.

Y

Decision principle.

Z

Judgment recommend allege.

Another fruitful investigation was that of Dr. W. Franklin Jones of the University of South Dakota. It differed somewhat in aim from that of Dr. Ayres. While Dr. Ayres dealt with the vocabularies of adults, Dr. Jones was concerned with the vocabularies of children only; and after tabulating as many as 15,000,000 words used in children's school compositions—a task which took him eight years—he published, among other things, a list of the 100 words most frequently misspelled by school children. He called them the "One Hundred Spelling Demons of the English Language." With the kind permission of Dr. Jones the Demons are printed on p. 246. The words are arranged in the order of the frequency with which the misspellings appeared, the most frequent being "which," the next "their," and so on down the columns in regular sequence.

TABLE VII—AYRES'S SPELLING SCALE

—	St. I.	St. II.	St. III.	St. IV.	St. V.	St. VI.	St. VII.
A . .	99						
B . .	98						
C . .	96	100					
D . .	94	99					
E . .	92	98					
F . .	88	96	100				
G . .	84	94	99				
H . .	79	92	98	100			
I . .	73	88	96	99			
J . .	66	84	94	98	100		
K . .	58	79	92	96	99		
L .	50	73	88	94	98	100	
M .		66	84	92	96	99	
N . .		58	79	88	94	98	100
O . .		50	73	84	92	96	99
P . .			66	79	88	94	98
Q . .			58	73	84	92	96
R . .			50	66	79	88	94
S .				58	73	84	92
T . .				50	66	79	88
U . .					58	73	84
V . .					50	66	79
W .						58	73
X . .						50	66
Y . .							58
Z . .							50

JONES'S HUNDRED SPELLING DEMONS

which	can't	guess	they
their	sure	says	half
there	loose	having	break
separate	lose	just	buy
don't	Wednesday	doctor	again
meant	country	whether	very
business	February	believe	none
many	know	knew	week
friend	could	laid	often
some	seems	tear	whole
been	Tuesday	choose	won't
since	wear	tired	cough
used	answer	grammar	piece
always	two	minute	raise
where	too	any	ache
women	ready	much	read
done	forty	beginning	said
hear	hour	blue	hoarse
here	trouble	though	shoes
write	among	coming	to-night
writing	busy	early	wrote
heard	built	instead	enough
does	colour	easy	truly
once	making	through	sugar
would	dear	every	straight

Let us now turn our attention from errors of spelling to errors in composition. About a year ago sixty head teachers of London elementary schools agreed to join me in a research into the nature of the errors made by children in their composition exercises. One exercise a month written

by each boy in the upper half of the school was carefully scrutinised, and the errors classified under the following heads :

1. Errors due to similarity of sound : *e.g.* is, his ; as, has ; to, too ; their, there ; were, where.
2. Misuse of "done" : *e.g.* He done it.
3. Misuse of other participles : *e.g.* seen for saw, come for came, etc.
4. Double negative : *e.g.* He didn't do nothing.
5. Double comparative : *e.g.* Richard is more bigger than Tom.
6. Failure of verb to agree with its subject : *e.g.* Everybody were pleased.
7. Wrong tense : *e.g.* He said that his father is ill.
8. Confusion of lie and lay, and of rise and raise.
9. Adjective used for adverb : *e.g.* He sang beautiful.
10. Wrong case of pronoun : *e.g.* Between you and I. Who did you see ?
11. Omission of full stop at end of sentence.
12. Omission of inverted commas.
13. Omission of question mark.
14. Omission or misuse of apostrophe : *e.g.* Dicks dog has lost it's tail.
15. Miscellaneous errors.

Altogether about 13,800,000 words were written and examined, and the errors detected therein amounted to 155,000 ; that is to say, a mistake of some kind appeared on an average once every 89 words.

The percentage of errors under each head,

estimated to the nearest half per cent., worked out as follows :

(1) 14·5, (2) 1, (3) 2, (4) ·5, (5) ·5, (6) 7, (7) 7·5, (8) 1, (9) 1·5, (10) 1·5, (11) 21, (12) 8·5, (13) 4, (14) 11, (15) 18·5.

The most frequent type of mistake included in the miscellaneous group was the omission or the misuse of capital letters. Clearly there ought to have been a separate head for this.

In an extensive investigation of this kind where the marking was done by different people with inevitably different standards, and where some gave concurrent instruction for the avoidance of the solecisms in question, while others delayed such instruction until the investigation was over, too much importance must not be attached to the exact aggregate of blunders. The distribution of blunders, however, among the various categories would not be seriously disturbed by differences of method and differences of standard. Indeed the results from the several schools were so similar in their general characteristics that the percentages given above may be accepted as reasonably trustworthy. And the following facts are worth pondering. Over a third of the errors come under two heads, 1 and 11, the first a kind of spelling error, and the second a kind of punctuation error. Nearly half the errors come under three heads, 1, 11, and 14. More than two-thirds of the errors come under six heads, 1, 6, 7, 11, 12, and 14. Punctuation errors cover 44 per cent. of the whole. If we leave out spelling and punctuation errors, only 21 per cent. are left as grammatical errors proper. This is at

variance with the view sometimes expressed, that because grammar receives but little attention in the elementary school, the essays written there abound in grammatical errors. One grammatical error in every 420 words written cannot be called alarming.

It is a significant fact that the greatest weakness revealed in this record lies in the pupils' loose grasp of the sentence as a unit of linguistic structure. In applying the remedy care should be taken that it is better than the disease. If a child gain a sense of structure and lose his impulse to write, his loss will be greater than his gain. So with all the other types of error: it is well to eliminate them, and to eliminate them with method and system; but we must not get so intent on weeding the garden as to forget to plant the seed and water the flowers. The casting out of errors is a necessary evil, and it is the main point and purpose of this chapter to prevent this necessary evil from becoming an unnecessary nuisance.

INDEX